mccsis 2015

multi conference on computer science and information systems

Las Palmas de Gran Canaria, SPAIN

21-24 July

Proceedings of the International Conferences

- Interfaces and Human Computer Interaction 2015,
- Game and Entertainment Technologies 2015 and
- Computer Graphics, Visualization, Computer Vision and Image Processing 2015

Edited by:
Katherine Blashki
Yingcai Xiao

ISBN 978-989-8533-38-8

iadis
international association for development of the information society

INTERNATIONAL CONFERENCES ON

INTERFACES AND HUMAN COMPUTER INTERACTION 2015

GAME AND ENTERTAINMENT TECHNOLOGIES 2015

and

COMPUTER GRAPHICS, VISUALIZATION, COMPUTER VISION AND IMAGE PROCESSING 2015

part of the

MULTI CONFERENCE ON COMPUTER SCIENCE AND

INFORMATION SYSTEMS 2015

PROCEEDINGS OF THE INTERNATIONAL CONFERENCES ON

INTERFACES AND HUMAN COMPUTER INTERACTION 2015

GAME AND ENTERTAINMENT TECHNOLOGIES 2015

and

COMPUTER GRAPHICS, VISUALIZATION, COMPUTER VISION AND IMAGE PROCESSING 2015

Las Palmas de Gran Canaria, Spain
JULY 22 - 24, 2015

Organised by
IADIS
International Association for Development of the Information Society

Co-organised by

UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Volume Editor:
Katherine Blashki
Yingcai Xiao

Computer Science and Information Systems Series Editors:
Piet Kommers, Pedro Isaías and Heredina Fernandez Betancort

Associate Editor: Luís Rodrigues

ISBN: 978-989-8533-38-8

TABLE OF CONTENTS

SHORT PAPERS

INTERFACES AND HUMAN COMPUTER INTERACTION

GAME AND ENTERTAINMENT TECHNOLOGIES

DOCTORALS

FOREWORD

These proceedings contain the papers of the International Conferences on Interfaces and Human Computer Interaction 2015, Game and Entertainment Technologies 2015 and Computer Graphics, Visualization, Computer Vision and Image Processing 2015 which were organised by the International Association for Development of the Information Society and and co-organized by Universidad de Las Palmas de Gran Canaria, Spain, 22–24 July, 2015. These conferences are part of the Multi Conference on Computer Science and Information Systems 2015, 21-24 July, which had a total of 652 submissions.

The Interfaces and Human Computer Interaction (IHCI) 2015 conference aims to address the main issues of concern within Interface Culture and Design with a particular emphasis on the affective aspects of design, development and implementation of interfaces and the generational implications for design of human and technology interaction. This conference aims to explore and discuss innovative studies of technology and its application in interfaces and welcomes research in progress, case studies, practical demonstrations and workshops in addition to the traditional submission categories.

This conference seeks to cover both technological as well as non-technological issues related to these developments.

Submissions were accepted under the following topics:
- Affective User-Centred Analysis, Design and Evaluation
- The Value of Affective Interfaces / Systems / Application / Interaction
- Generational differences and technology design
- Measurement of success of emotional technology / interfaces
- Supporting user populations from specific Generations
- Supporting user populations with Physical Disabilities
- Supporting user populations with Intellectual Disabilities
- Creativity Support Systems
- Emotional Design issues / methods / experiences for novel interfaces including tangible, Mobile and Ubiquitous computing, mixed reality interfaces and multi-modal interfaces
- Usability
- User studies and fieldwork
- Methodological implications of Emotional User Studies
- Participatory design and Cooperative design techniques
- Ethical issues in emotional design
- HCI and Design education
- Eliciting User Requirements

The Game and Entertainment Technologies (GET) 2015 conference aims to bring together research and practice from creative, social and business practitioners and researchers in this challenging field. The focus of this conference is on design, development and evaluation of games, entertainment technologies and the nature of play.

Known to have been enjoyed since at least 30 BC, games and entertainment are a universal part of human experience and present in all cultures. Games and entertainment activities contribute to the social, emotional, psychological and physical well-being of human society. As game and entertainment technologies become increasingly more pervasive we are continually challenged in our work, learning and personal life by increased access to virtual spaces and communities that offer opportunities for everyday needs and aesthetic experiences. The 'Creative Industries' require design and development structures, techniques and methodologies that enrich, enhance and encourage new interaction modes, metaphors and in-depth co-creation.

Topics of interest include, but are not limited to the following areas:

- Development methodologies
- Design issues
- Controversial issues – we welcome debate and dissension, for example; games as art, entertainment as purely for monetary returns etc
- Special Effects
- Animation
- Mobile and ubiquitous games and entertainment
- Serious Games –applications, critiques
- Philosophical issues
- Prototypes
- Social and cultural uses of/for Play
- Tools and technologies
- Skills, strategy, rules and chance
- Genre
- Immersiveness and engagement
- Research methodologies in creative practice
- Usability and playability
- User/player centered design
- Psychological, social, and cultural differences in perception and participation
- Communities, networks, social interaction and social capital
- Cross-cultural and intercultural approaches
- Assessment of exploratory learning approaches
- Emerging practices

The Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2015 conference intends to address the research issues in the closely related areas of Computer Graphics, Visualization, Computer Vision and Image Processing. The conference encourages the interdisciplinary research and applications of these areas.

Submissions were accepted under the following 5 main topics:

- Computer Graphics
- Visualization
- Computer Vision
- Image Processing
- Other Related Topics

These events received 178 submissions from more than 27 countries. Each submission has been anonymously reviewed by an average of five independent reviewers, to ensure that accepted submissions were of a high standard. Consequently only 28 full papers were approved which means an acceptance rate of 16%. A few more papers were accepted as short papers, reflection papers, posters and doctoral papers. An extended version of the best papers will be published in the IADIS International Journal on Computer Science and Information Systems (ISSN: 1646-3692) and/or in the IADIS International Journal on WWW/Internet (ISSN: 1645-7641) and also in other selected journals, including journals from Inderscience.

Besides the presentation of full papers, short papers, reflection papers, posters and a doctoral consortium the conferences also included one keynote presentation from an internationally distinguished researcher. We would therefore like to express our gratitude to Professor Joaquim Jorge, Full Professor, Department of Computer Science and Engineering, IST - Técnico Lisboa, Portugal for accepting our invitation as keynote speaker. Also a special thanks to Nitendra Rajput, Senior Researcher & Research Manager, IBM Research, India for is presence as an invited speaker.

This volume has taken shape as a result of the contributions from a number of individuals. We are grateful to all authors who have submitted their papers to enrich these conferences proceedings. We wish to thank all members of the organizing committee, delegates, invitees and guests whose contribution and involvement are crucial for the success of the conference.

Last but not the least, we hope that everybody will have a good time in Las Palmas de Gran Canaria, and we invite all participants for the next edition.

Katherine Blashki, Noroff University College, Norway
Interfaces and Human Computer Interaction 2015 Program Chair
Game and Entertainment Technologies 2015 Program Chair

Yingcai Xiao, The University of Akron, USA
Computer Graphics, Visualization, Computer Vision and Image Processing 2015 Program Chair

Piet Kommers, University of Twente, The Netherlands
Pedro Isaías, Universidade Aberta (Portuguese Open University), Portugal
Heredina Fernandez Betancort, Universidad de Las Palmas de Gran Canaria, Spain
MCCSIS 2015 General Conference Co-Chairs

Las Palmas de Gran Canaria, Spain
July 2015

PROGRAM COMMITTEE

INTERFACES AND HUMAN COMPUTER INTERACTION CONFERENCE

GAME AND ENTERTAINMENT TECHNOLOGIES

PROGRAM CHAIR

Katherine Blashki, Noroff University College, Norway

COMPUTER GRAPHICS, VISUALIZATION, COMPUTER VISION AND IMAGE PROCESSING

PROGRAM CHAIR

Yingcai Xiao, The University of Akron, USA

MCCSIS GENERAL CONFERENCE CO-CHAIRS

Piet Kommers, University of Twente, The Netherlands
Pedro Isaías, Universidade Aberta (Portuguese Open University), Portugal
Heredina Fernandez Betancort, Universidad de Las Palmas de Gran Canaria, Spain

INTERFACES AND HUMAN COMPUTER INTERACTION CONFERENCE

COMMITTEE MEMBERS

Aidan Rowe, University of Alberta, Canada
Alessandro Casavola, University of Calabria, Italy
Alexander Gelbukh, Instituto Politécnico Nacional, Mexico
Alexander Kröner, Technische Hochschule Nürnberg Georg Simon Ohm, Germany
Ana Cristina Bicharra Garcia , Universidade Federal Flumninese, Brazil
Ant Ozok, UMBC, USA
Antonio Maria Rinaldi, University Napoli Federico Ii, Italy
Arianna D Ulizia, Irpps-cnr, Italy
Arttu Perttula, Tampere University of Technology, Finland
Bruno Santana Da Silva, Federal University of Rio Grande do Norte, Brazil
Caterina Senette, Istituto Di Informatica e Telematica, CNR (pisa), Italy
Celia Martinie, Universite Paul Sabatier, France
Connie Veugen, Vrije Universiteit Amsterdam, The Netherlands
Corina Sas, Lancaster University, United Kingdom
Daniela Trevisan, Universidade Federal Fluminense, Brazil
Eduardo Tanaka, Universidade Estadual de Campinas, Brazil
Elaine Hayashi, University Of Campinas, Brazil
Evan Golub, University Of Maryland, USA
Fazel Naghdy, University Of Wollongong, Australia
Florian Reinfrank, Tu Graz, Austria
Francisco Jurado, Universidad Autónoma de Madrid, Spain

Toni Granollers, University of Lleida, Spain
Vagner Figueredo de Santana, IBM Research, Brazil
William Joseph Giraldo Orozco, Universidad del Quindío, Colombia
Xiaokun Zhang, Athabasca University, Canada
Yacine Bellik, Laboratory Limsi-cnrs, France

GAME AND ENTERTAINMENT TECHNOLOGIES

COMMITTEE MEMBERS

Abdennour El Rhalibi, Liverpool John Moores University, United Kingdom
Alf Inge Wang, Norwegian University Of Science and Technology, Norway
Ali Arya, Carleton University, Canada
Anastasios Karakostas, Aristotle University of Thessaloniki, Greece
Anthony Whitehead, Carleton University, Canada
Areti Damala, University of Strathclyde, UK
Armelle Prigent, Universite de la Rochelle/L3I, France
Arttu Perttula, Tampere University of Technology, Finland
Brunhild Bushoff, sagasnet, Germany
Carla Hoekendijk, Independant Consultant/Developer, The Netherlands
Castulus Kolo, Macromedia University of Applied Sciences, Germany
Christos Bouras, University of Patras and Computer Technology Insti, Greece
Connie Veugen, Vrije Universiteit Amsterdam, The Netherlands
Daniel Cermak-sassenrath, IT University of Copenhagen, Denmark
Ellen Brox, Norut – Northen Research Institute, Norway
Emanuele Carlini, ISTI, CNR, Italy
Emmanuel Tsekleves, Lancaster University, United Kingdom
Eric Gressier, Cnam – Cedric, France
Eva Hudlicka, Psychometrix Associates, USA
Gabriele D`Angelo, University of Bologna, Italy
Gunver Majgaard, University of Southern Denmark, Denmark
Harry Agius, Brunel University, United Kingdom
Hassan Qudrat-Ullah, York University, Canada
Ian Marshall, Coventry University, United Kingdom
Ioannis Paraskevopoulos, Anglia Ruskin University, United Kingdom
Isabelle Astic, CNAM, France
Jacco Bikker, NHTV University of Applied Science, The Netherlands
James Bowman, University of Advancing Technology, USA
Jerome Dupire, Cnam – Cedric, France
Jon Preston, Southern Polytechnic State University, USA
Jose Castro-filho, UFC-virtual, Brazil
Jouni Smed, University of Turku, Finland
Kai Kimppa, Turku School of Economics, Finland
Laura Ricci, University of Pisa, Italy
Liselotte Van Leeuwen, University of Sunderland, United Kingdom
Magnus Johansson, Stockholm University, Sweden
Marcia Medeiros, UFC – FANOR (Devry), Brazil
Marcos Rodrigues, Sheffield Hallam University, United Kingdom
Markus Wiemker, Rwth Aachen, Germany

Melinda Jacobs, Subatomic, The Netherlands
Michael Katchabaw, University of Western Ontario, Canada
Michel Simatic, Telecom Sudparis, France
Mikael Collan, Lappeenranta University of Technology, Finland
Monica Gavrielidou, University of Thessaly, Greece
Nicolas Auray, Telecom Paristech, France
Niklas Ravaja, University of Helsinki, Finland
Pauliina Tuomi, Tampere University of Technology, Finland
Pilar Lacasa, University of Alcala, Spain
Richard Bartle, University of Essex, United Kingdom
Robert Stone, The University of Birmingham, United Kingdom
Rommert Casimir, Tilburg University, The Netherlands
Ruck Thawonmas, Ritsumeikan University, Japan
Sarah Labelle, Université Paris XIII, France
Spyros Vosinakis, University of the Aegean, Greece
Stefano Ferretti, University Of Bologna, Italy
Stephane Natkin, CNAM, France
Steven Harris, Anglia Ruskin University, United Kingdom
Thrasyvoulos Tsiatsos, Aristotle Univerisity of Thessaloniki, Greece
Timo Lainema, Turku School of Economics, University of Turku, Finland
Yoshihiro Okada, Kyushu University, Japan

COMPUTER GRAPHICS, VISUALIZATION, COMPUTER VISION AND IMAGE PROCESSING

COMMITTEE MEMBERS

Adrian Jarabo, University of Zaragoza, Spain
Aiert Amundarain, CEIT, Spain
Alberto Raposo, PUC-rio, Brazil
Alessandro Artusi, University of Girona, Spain
Alessandro Rizzi, Università Degli Studi di Milano, Italy
Andre Hinkenjann, Institute of Visual Computing, Bonn-Rhein-Sieg Uni, Germany
Angelica De Antonio, Universidad Politecnica de Madrid, Spain
Arcadio Reyes Lecuona, Universidad de Málaga, Spain
Belen Masia, Universidad de Zaragoza, Spain
Branka Medved Rogina, Institute Ruđer Bošković, Croatia
Bruce Campbell, Rhode Island School of Design, USA
C. C. Lu, Kent State University, USA
Carla Binucci, Università Degli Studi di Perugia, Italy
Carlo Nati, Working Group for development of Scientific and Technological Knowledge Italian Ministry of Education, Italy
Celmar Guimarães da Silva, University of Campinas, Brazil
Cesar Alberto Collazos, Universidad del Cauca, Colombia
Chang Ha Lee, Chung-ang University, Republic of Korea
Charalambos Poullis, Cyprus University of Technology, Cyprus
Charalampos Georgiadis, The Aristotle University, Greece

Luciano Soares, Insper Instituto de Ensino e Pesquisa, Brazil
Luis Molina-Tanco, Universidad de Malaga, Spain
Luis Unzueta, Vicomtech-IK4, Spain
M. Emre Celebi, Louisiana State University in Shreveport, USA
Marc Daniel, Ecole Superieure D ingenieurs de Luminy, France
Marcelo Guimaraes, UNIFESP / FACCAMP, Brazil
Mubbasir Kapadia, Disney Research Zurich, Switzerland
Mu-Chun Su, National Central University, Taiwan
Nadia Ambrosetti, Universita Degli Studi Di Milano, Italy
Nicoletta Adamo-villani, Purdue University, USA
Olga Sourina, Nanyang Technological University, Singapore
Paolo Pingi, CNR – ISTI, Italy
Pascual González López, University of Castilla-La Mancha, Spain
Patrick Gioia, Orange Labs, France
Pedro Cano, Universidad de Granada, Spain
Peter Dannenmann, RheinMain University of Applied Sciences, Germany
Piotr Lech, West Pomeranian University of Technology, Poland
Quan Wen, University of Electronic Science and Technology of China, China
Rene Rosenbaum, University of Rostock, Germany
Robert Laramee, Swansea University, United Kingdom
Roman Durikovic, Comenius University in Bratislava, Slovakia
Sandra Morales, Polytechnic University of Valencia, Spain
Shaun Bangay, Deakin University, Australia
Simon Richir, Arts et Metiers ParisTech, France
Stanislav Klimenko, Moscow Institute of Physics and Technology (state), Russia
Stella Sylaiou, Aristotle University of Thessaloniki, Greece
Stratos Stylianidis, Aristotle University of Thessaloniki, Greece
Sung-Hee Lee, KAIST, South Korea
Theresa-marie Rhyne, Computer Graphics Consultant, USA
Tim Newman, University of Alabama in Huntsville, USA
Valery Adzhiev, Bournemouth University, United Kingdom
Victor Debelov, Institute of Computational Mathematics and Mathematical Geophisycs of SB RAS, Russia
Vladimir Savchenko, Hosei University, Japan
Wojciech Wiza, Poznan University of Economics, Poland
Xenophon Zabulis, FORTH, Greece
Xiaogang Jin, Zhejiang University, China
Xiaosong Yang, Bournemouth University, United Kingdom
Xiaoyue Jiang, Northwestern Polytechnical University, China
Xuejun Hao, Columbia University, USA
Zhong-hui Duan, University of Akron, USA
Zongyi Liu, Microsoft, USA
Zoran Ivanovski, Faculty of Electrical Engineering and Information, Macedonia

KEYNOTE LECTURE

GAME OVER? NEW APPROACHES TO TEACHING ENGINEERING COURSES

Professor Joaquim Jorge
Department of Computer Science and Engineering
IST - Técnico Lisboa, Portugal

ABSTRACT

Gamification has been explored recently as a way to promote content delivery in education, yielding promising results. However, little is known regarding how it helps different students experience learning and acquire knowledge.

In this talk I discuss our experiences with gamified engineering courses and the reactions of students to the gamified experience.

By examining student performance and attitude data collected from several years we identified distinct student types. I will describe different student types, according to behavior and explain how gamification can provide for smarter learning by catering to students with different profiles.

INVITED SPEECH

"HCI FOR THE NEXT BILLIONS"

Nitendra Rajput, Senior Researcher & Research Manager,
IBM Research, India

ABSTRACT

Every user likes easier interfaces. But for people who have physical, intellectual, literacy or monitory challenges, an easier interface makes or breaks the underlying application or solution. This talk will provide insights into developing HCI systems for people with special needs. Specifically, it will address design of speech based interfaces for low-literate users, design of gesture based systems for people with intellectual challenges, design of haptic mobile based systems for technologically challenged users. These gamut of examples will highlight the importance of HCI for the otherwise unaddressed community. The talk will also highlight the potential benefits associated with designing innovative interactions for such communities.

Full Papers

HCI AND DESIGN THINKING: EFFECTS ON INNOVATION IN THE ACADEMIC LIBRARY

Alma L. Culén[1] and Andrea A. Gasparini[2]
[1]*Department of Informatics, University of Oslo, P. Boks 1080, 0316 Blindern, Oslo, Norway*
[2]*Department of Informatics and Oslo University Library, University of Oslo, P. Boks 1080, 0316 Blindern, Oslo, Norway*

ABSTRACT

This paper is concerned with how two design processes that are seemingly very similar, human-computer interaction design and design thinking, affect innovation in the context of an academic library. Twenty different multidisciplinary projects involving advanced students of interaction design, researchers and library employees, using either of these two approaches or their combination, were conducted during the past four years. We chose two projects that we consider as illustrative of overall findings, in particular, regarding outcomes for diverse stakeholders. Both projects were focusing on the same problem, which many academic (and public) libraries face, of accessing library's e-book collections and making doing so more enjoyable.

KEYWORDS

Design thinking, human-computer interaction, innovation, user experience, leap-motion, apps, e-books.

1. INTRODUCTION

For the past two decades, the Internet has been a game-changer for academic libraries. Appearance of disruptive technologies, such as e-books first, and tablets later, posed further challenges. Also, fundamental changes in users' behavior became apparent. Users were increasingly demanding regarding the quality of services that they use, as well as capable and creative in their use of technology. Following users and their technology use patterns became a problem for many traditional institutions, among them also academic libraries. Many libraries worldwide are turning to service and user experience (UX) design, and design labs to address the problem (Mathews, 2012; Rundblad, 2011; Schmidt and Etches, 2012). Academic libraries are particularly hard pressed to re-think their role in academic life, technology use and willingness to innovate.

Four years ago, open user-driven innovation (Von Hippel, 2009; Chesbrough, 2007) was chosen as an approach to create innovative solutions that improve existing, or introduce new services and user experiences in the context of the academic library. The innovators were users, who also were students of human-computer interaction (HCI), and the innovation processes were carried out as semester-long projects within an HCI course. Even though that year no real innovation took place (in the sense of producing new products or services that were made available for use in the library), the student projects were evaluated positively by the library. The approach seemed promising since it brought forth users' perspective on library services. Consequently, a long-term cooperation was established between the library and interaction design students and researchers. Over the next two years, additional projects (and resources from the library) were dedicated to developing innovative interfaces and services through this cooperation that was then re-framed as a living lab. The living lab framework implied a strong focus on users, innovation, and co-creation (Bergvall-Kåreborn and Ståhlbröst, 2009).

This paper describes what we learned regarding differences between design approaches – the HCI design and the design thinking – in supporting user-centered innovation and co-creation processes. In particular, we were interested in finding out how the choice of the approach affected stakeholders. As researchers associated with the living lab, we have done participatory observations, and interviews with students and library employees after each project. Our main finding is that both approaches support innovation, but in different ways, and with different effects on stakeholders, as this paper aims to show.

Two projects, one using HCI design and the other both HCI and design thinking combined, are chosen to illustrate our overall findings. Both projects explored the problem of finding e-books, and making people aware of their existence. Many academic libraries are trying to solve this problem (Buczynski, 2010; Shelburne, 2009), and find it to be complex. The complexity stems from the combination of human causes (e.g., users actually not knowing what an e-book is, or not recognizing it even when looking at it) and technical issues (e.g., systems and platforms for finding e-books that are unintentionally hidden from users).

The paper is structured as follows: in the next section, we present the background for this paper, followed by a section presenting the two interface design cases, solving the e-book finding and awareness problem using either HCI or design thinking, in combination with HCI, approaches. The discussion section then focuses on the effect of HCI and design thinking approaches on partners/stakeholders and further activities in the living lab. Conclusion closes the paper.

2. BACKGROUND

We start by establishing a common understanding of what an HCI design process is, at least in the academic context (HCI practitioners in the industry may have a different understanding). Even within academia, there are quite a few opinions on the theme. However, we believe that we take the mainstream approach when describing HCI design processes as activities starting from understanding the design context and identifying the needs and requirements for design, creating a conceptual design, implementing the solution (prototyping), evaluating it and then iterating (including new requirements, if found, or redesigning based on results of evaluations, and repeating these steps) (Sharp et al., 2007, p. 448).

Research through design, advocating design practices as a way of supporting the research in HCI, gained ground in HCI through works such as (Fallman, 2003; Forlizzi et al., 2008; Zimmerman and Forlizzi, 2014). The difference between the traditional HCI approach and research through design is that the latter brings design (and design methods from disciplines engaged with formal design training, such as product design, art or architecture) and designerly practices into HCI.

For the passed two decades, design and designers have, according to (Krippendorff, 2005), taken a semantic turn, seeking to embed meaning in what they do in the world and take a lead in solving complex problems. Design thinking (Brown, 2009) aided this movement, and established itself as one of the drivers of innovation and competitive advantage in business (Culén and Kriger, 2014; Lockwood, 2009; Martin, 2009). According to (Martin, 2009), everyone can work on becoming a design thinker. In order to do so, one needs a stance, tools, and experience that facilitate design thinking. The stance is related to one's worldview and the role one has in it, tools are the mental models used to understand the world and organize thinking, while experience is needed for building of skills and sensitivities. This implies that one becomes a keen observer and finder of opportunities for design (Wagner and Compton, 2012). The three main pillars of design thinking are its user-centered approach, rapid prototyping, and abductive thinking, or the inference to the best explanation. The latter is often quoted as a necessary tool design thinkers should have (Kolko, 2012). In spite of the fact that theoretically, everyone can become a design thinker, what we see in the literature are predominantly success stories of how design thinking processes give good results when guided by professional designers, e.g., (Brown, 2009, 2008). Papers that present how to establish well-working, novice multidisciplinary teams using design thinking (Seidel and Fixson, 2013), are more seldom. The design thinking processes typically include understanding of the problem space and empathy with users, framing the problem, divergent thinking, creative thinking, generating many ideas and possible solutions, using abduction to find best options (convergent thinking, sensibilities), prototyping, testing, and iterating, similar to HCI.

While the two approaches, HCI design and design thinking, may be seemingly similar, their effects often differ. HCI design processes usually involve incremental innovation and small changes (e.g., new versions of existing software), that are readily supported by the iterative nature of the process (Culén and Følstad, 2014). Design thinking, on the other hand, offers a possibility for radical innovation through divergent and convergent thinking (Runco and Acar, 2012) and focus on proper framing of the problem (e.g., rather than solving a problem, solve the right problem). Also, while HCI design processes rarely bring about organizational changes, design thinking approach to innovation often affects the whole organization (Brown, 2009, 2008; Leavy, 2010; Martin, 2009).

The idea that design thinking could be integrated in HCI education led to re-examining the teaching of HCI (Culén, 2015; Culén et al., 2014). It was found that one of the main obstacles to successfully integrating two practices is students' perception that design and design practices require talent, or the above discussed abilities such as abductive thinking and designers' sensibilities. HCI students (prior to the exposure to creative practices) often stated that they are not creative or, that they feel more comfortable following prescribed procedures rather than more open processes. However, after being exposed to creative thinking techniques, they felt, and reported so through the end of the year questionnaire, that they became more confident, learned and grew professionally in the process (Culén, 2015). We now present the two illustrative design projects, both engaging computer science students in solving the same problem of finding e-books.

3. CASE: FIND E-BOOKS!

These days, our academic library, in likeness with many others, uses the majority of its funds on digital content. Yet, this content is either not readily available to users or is so seamless that users do not realize that they are actually using the library resources when, for example, downloading articles that are free of charge to them because the library already obtained rights to the content. The now large quantity of library's e-books remains underused by the academic community. This is in part due, the library believes, to the fact that they have no good solutions for making people aware that most books that they view as paper books in the library, also exist in the digital edition. What is easy on Amazon, often gets forgotten in the academic library.

As mentioned in the introduction, the library's living lab innovation efforts are implemented through co-creation between students from a project-based interaction design course at the computer science department, researchers and library employees. As described in (Culén, 2015), the course is a combined bachelor-master level, teaching research methods in HCI (Lazar et al., 2010). Two prior HCI courses are prerequisites. About half of the fifteen weeks (duration of the course) is used for concept development and prototyping, and the other half for evaluation and experiment design, including both qualitative and quantitative analysis methods.

We now describe the first project, Bookworms (Reistad et al., 2012) from 2012, which followed a traditional HCI design approach and was part of the user-driven innovation approach. The project team consisted of three library employees, two researchers, and five HCI design students (as innovators). The idea that a mobile phone app could help find a book in the library building motivated the students highly. Apps were cool, and the team settled on this idea from the very beginning. They then worked hard to implement the app through iterative improvements, until the product was finished and launched through both Google Play and Apple Store ("Realfagsbiblioteket-Google," 2013; "Realfagsbiblioteket-Apple," 2013). The app is in use today, in its third version, surprisingly, also by the librarians.

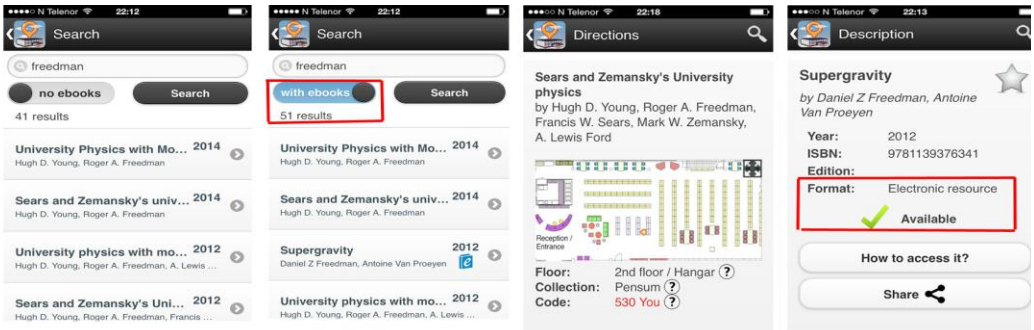


Figure 1. Screenshots of the app, with search without e-books for Freedman's books, and with e-books, followed by a screen showing how to find a book in the library, and a screen showing whether an e-book is available, and if so, how to get it.

As part of their overall solution for finding books in the library, an easy way to check whether a book is available in digital edition or not, was provided. As shown in Figure 1, the application's main search had an e-book option, and provided a way of accessing e-books once they were found. The app was well designed, with a clean user interface. The library helped with relevant information on their current search systems and databases. The prototype made for the course was fully functional but used a small portion of the book collection. The students were consequently hired to finish the work for the entire science library. The

ISBN: 978-989-8533-38-8 © 2015

students also wrote a research report, describing their methods for gathering data, ways of involving users in the design, and research results related to the usability testing (Reistad et al., 2012).

In preparation for the fall 2013 projects, where the library wished to continue looking with solutions to e-book collections presentations (online and in the library), a workshop was conducted, introducing design thinking and relevant service design concepts to library employees. The workshop had 25 participants. Seventeen participants were library employees, including some leaders, librarians, digital services management and support, e-resources and open access consultants, and others. Four participants were graduate students in interaction design and four were researchers. The participants were divided into four multi-disciplinary groups. The outcome of the workshop was, for the library employees, the understanding that service design uses a *design thinking approach* and tools such as *customer journeys* and *touch points*. A customer journey is the complete sum of experiences that users have when interacting with the library. Touch points are contact points between users and the library (where understanding user position is important). They can be digital (e.g., a website), or physical (e.g., a front desk). Customer journeys (prototypes of future services) were easy to visualize using service design cards (touch points cards). Design of good future services depends on considering diverse journeys and infering the best solutions.

The workshop was hands-on. Each team got their deck of service design cards, dots and arrows, a large sheet of paper, and colored pens. The participants took some time to become familiar with cards. After that, the cards became an excellent tool, facilitating the building of common understanding of the problem space. Soon, all groups started also using arrows and dots to reason by assigning importance to specific cards or processes and mapping out present and future customer journeys while searching for e-books, see Figure 2. Design thinking was an approach, at least in the simplified format in which it was presented, that all could easily understand and use. A less visible, but perhaps more important, outcome of the workshop had to do with awareness building around the role that design thinking may play in library innovation processes, in line with what is claimed in (Brown, 2009). The workshop was evaluated as informative and engaging.

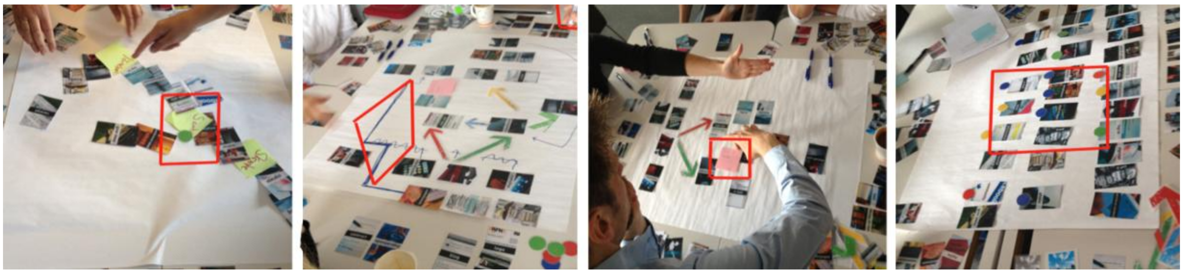


Figure 2. The workshop with library employees, students and researchers: creating customer journeys related to e-book searches, both present and future. Finding and considering importance of touch points.

The second design case, the project BookMotion (Okun et al., 2013), was carried out during the fall of 2013. Similarly to the first project described, it resulted in a high-fidelity prototype that is used also today.

As described, we have engaged the library employees in the workshop prior to the start of the course. In order to give students the same basic knowledge regarding design thinking, customer journeys and touch points, the workshop was repeated with students who chose this project. Now, all team participants (students, researchers and three to eight library employees that attended design sessions) had the same basic knowledge to work with. Rather than simply providing information, as was the case in the first process, the library employees were now much more active, clearly showing interest to participate as co-creators in the process. This interest, in combination with design thinking approach and new tools (service design cards, customer journeys and so on) motivated the students highly. A new technology, Leap Motion, appeared on the market just before the start of the course and offered new design opportunities. A virtual browse of a physical library was already tested and described in (Lynema et al., 2012). The team added the Leap Motion as a new touch point. Considering different possibilities for placing this touch point on customer journeys (rapid prototyping), they found out that, initially, the best solution for awareness building around e-books is for it to start within the library building. Bringing the virtual (e-books) into the library room was seen as interesting, and worthy of further exploration. The task now included more than prototyping with the Leap-motion, it also included the design of an interactive space in the library, see Figure 3, building awareness around what e-books are and how to get them. This was a bit more open and 'designerly' task than what was common for HCI students. (BookMotion by Y-TEK, 2013) shows a short video demonstration of the project.

Figure 3. Browsing gesture while searching for e-books. When an e-book is selected, the QR code is scanned and the link to the e-book sent to the user's email address. Placement of the system within the library (Okun et al., 2013).

However, in the prototyping phase, it became clear that the technology, the Leap Motion, had its limitations. The 'natural' gestures were not the same for all users, and finding the 'right' natural gestures to browse, select, and check out e-books turned out to be a problem that is not simple to solve. The Leap Motion had issues with interpreting the same gesture performed by different users accurately. The students used the HCI approach this time. Different gestures were carefully tested in order to find the most appropriate ones for the final prototype. Ways of displaying books, amounts of space between the books enabling the most comfortable and satisfactory interactions (the latter depending strongly on precision needed for correct selection), were tested as well. However, many users continued to experience frustration despite the best efforts from the team, and the problem remained unsolved.

As this project was carried out within the living lab framework, the research was done in–the–wild, at the library. In addition, as a biology conference took place during the project, it was used for testing purposes. Natural gestures-based search was implemented for a small collection of biology books that were new and being introduced to conference attendees. Many attendees tried the system, and despite above-mentioned problems, liked the interface, in particular, the newness of the interaction. The same conference recently made a request to the library to have the BookMotion installation available again. The library thus can re-use the code and with moderate investment of time, make the installation work for special occasions, such as this conference, where the number of books to search through is relatively limited.

4. DISCUSSION: EFFECTS ON INNOVATION PROCESSES

A panel of four professional interaction designers and HCI practitioners selected the Bookworm project as the best HCI student project for the year 2012. The resulting app was *playful* and had a *great functionality*. The research related to *usability and user satisfaction* was rigorous and the results sound. The most fun part of the app, finding a physical book on the library shelf easily, was location bound. With relatively small efforts and further investment, the app could be made to work in other libraries. The search for e-books, though, was not location bound. With this in mind, the next generation of interaction design students was offered a possibility to consider making an app that facilitates browsing of e-book collections, as its primary functionality. Students, however, felt that they did not have much to add, and no team chose this possibility. One way they described the sentiment was that the app was "too finished".

In contrast, the BookMotion project, which ended with a number of obvious problems, was picked up first by the library that developed interactions with the system further, adding a student card reader, enabling knowing the identity of a person downloading the e-book. A special collection, ("Collection 42 - University of Oslo Library," 2015) was designated for use with the BookMotion. Furthermore, the medical library wished to use system's gesture-based interactions to browse collections of human anatomy images.

A new student team starting a year later, joined in, utilizing both design thinking and agile programming to implement changes in the interface addressing gesture precision difficulties. The design thinking approach was visible in creative generation of rapid prototypes, hands-on (sensorial) exploration of intermediary surfaces such as a soft touch screens made of plastic foil, bubbles that need to be popped in order to make a book selection, diverse magic or conductor sticks, were tried. Some prototypes tried are shown in Figure 4.

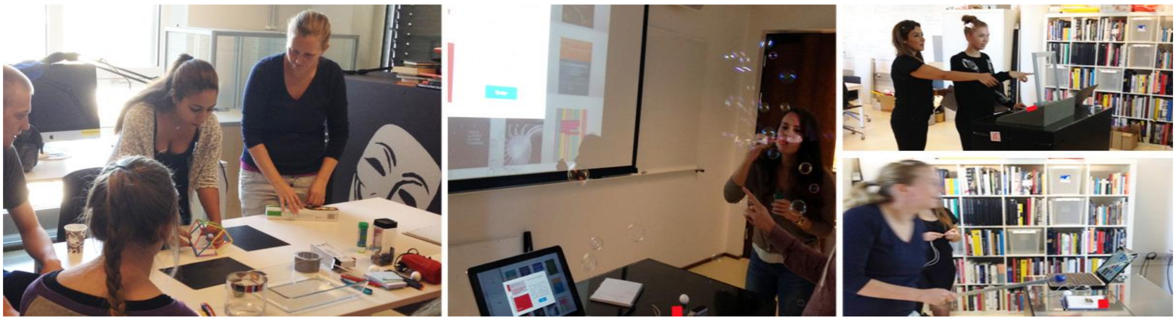


Figure 4. Co-creating an intermediate interface workshop (with Paria Tahaee, a master student in interaction design).

So why was there so much interest in BookMotion? *"It felt like a building block, like a Lego block, you just have to play",* said one member of the student team, referring to the Leap itself. This remark illustrates the general trend showing how *cool new technology* becomes an important actor in innovation processes, in line with what design thinking literature shows, e.g., (Brown, 2009). The *openness* of the original application, both with regard to the technical platform (open source on Github) and the possibility to build easily upon the basic applications were important to many students. Another student liked the *value* aspect of this project: *"As a developer, I get a good feeling when I develop not only for myself, but also for others. I thought that this LeapMotion is good for bringing awareness around e-books. After one understands what it is about, one can browse on all platforms. So, I think it is useful."* And finally, the power of design thinking was seen, in particular, through the power of divergent and convergent thinking.

The convergent thinking implies the understanding of the known in a new light, the ability to see new possibilities. It also implies taking risks (Cropley, 2006). The LeapMotion was a new technology at the time, and students clearly took a risk when choosing it as a platform for a public service. The openness of their solution triggered other students to try to find solutions to gesture recognition and precision problems, attempting prototypes such as the ones shown in Figure 4 and forcing the new group to take several types of risks. The uncertainty of the final result for the library was a certainty, as were possible negative implications on the students' course grade. In contrast to convergent thinking, divergent thinking invites to failure, and opportunity to learn from mistakes. This was also the case in this project, as many of the prototypes were not functioning or were difficult to implement.

This risk taking aspects and invitation to fail are not common for HCI design projects, where it is nearly always possible to improve existing solutions, or design new ones following steps that (nearly) insure success. In design thinking and innovation, risk taking is the integral part of the process. In addition, producing and using new knowledge is required (else, there is no innovation) (Cropley, 2006). When using designerly practices, the knowledge production mostly happens through making, through rapid prototyping. Within the co-creating teams, where participants brought with them different type of knowledge based on their experiences, we could see how new knowledge and perspectives were added through the process of making. The latter may be seen as construction design research, which refers to design research where the construction of a product, system, space, or media is central, and becomes the key means in producing new knowledge (Koskinen et al., 2011). On the other hand, the knowledge produced by HCI was often generated as in natural sciences, through hypothesis postulation and experimentation. The risk taking, failure, learning and knowledge production, or absence of those, were all clearly observable through innovation projects, whatever design approach was chosen.

Since making of high fidelity (working) prototypes was a course requirement, nearly all teams could try and test their solutions in realistic (living lab) settings. The prototypes have, as a rule, generated a lot of research questions. The nature of questions and answers differed between two approaches (HCI design and design thinking), one being more focused on answering 'why is this so?' and the other 'how can I make this work?' This is illustrated by the first attempt to solve the gesture recognition problem for BookMotion through HCI research, and the second, trying to make it work by circumventing the problem (Figure 4).

From students' questionnaires, filled at the end of the year, all student teams that used design thinking (alone, or in combination with HCI design processes) reported that they now consider creativity to be important, and think that what they do is creative. In addition, they find the group work, particularly in truly multidisciplinary teams, to be of great importance. Contributions from diverse members created a knowledge base that, in addition to creativity, significantly increased chances of creating good innovative solutions. We believe that for students, this may be the most important demonstration of what good learning practices are.

5. CONCLUSION

It follows from the discussion above, as well as from our observations of other projects, that cool new technology, openness, values and design thinking (in particular divergent and convergent thinking) make a powerful mix for innovation purposes in the context of the living lab and co-creation with interaction design students, researchers and library employees. This finding is in line with literature on open innovation and design thinking.

Also the students who tried innovation in the traditional HCI way, have achieved remarkable results. They tended to produce functional products, giving good user experience, and their evaluation and analysis were good, as the Bookworm project shows.

The interesting new insight is that we believe that HCI education should increase focus on creative thinking and integrate it more profoundly in starting phases of HCI design processes, when novelty and innovation are desired. Based on students feedback, as well as participatory observations, we further conclude that constructive design research and designerly practices, when focusing on innovation, provide for good learning practices for HCI education. Since computer science students are not trained designers, designerly practices and design thinking may be opening the door to creativity in appropriate ways and in line with todays increasingly accepted constuctivist learning practices.

In relation to the library, some innovative products were made during the past four years. These, when taken in use, as Bookworms and BookMotion, could be called innovations. However, the most important changes for the library and the library employees have been the organizational changes influenced by design thinking. Gaining the confidence to move from information providers and observers, as in Bookworms project, the library employees embraced the design thinking and the possibility to co-create, as in BookMotion, making them (co-)own ideas and products. That opened room for design-driven (or, more correctly in our case, designerly-driven) approach to innovation, and made the library into a proper living lab.

ACKNOWLEDGEMENT

This work was, in part, supported by User-Driven Innovation and Library UX projects. Thanks are due to all library employees and student teams who participated in the project.

REFERENCES

Bergvall-Kåreborn, B., Ståhlbröst, A., 2009. Living Lab: an open and citizen-centric approach for innovation. *International Journal of Innovation and Regional Development*, 1(4), 356–370.

BookMotion by Y-TEK, Youtube video, 2013. Available at https://www.youtube.com/watch?v=OP_-zJIcJRY.

Brown, T., 2009. *Change by design: how design thinking can transform organizations and inspire innovation.* Harper Collins Publishers, New York, NY.

Brown, T., 2008. Design Thinking. *Harvard Business Review*, Vol. 86, No. 6, pp. 84-92.

Buczynski, J. A. 2010. Library eBooks: Some Can't Find Them, Others Find Them and Don't Know What They Are. *Internet Reference Services Quarterly*, Vol. 15, No. 1, pp. 11–19.

Chesbrough, C., 2007. Why Companies Should Have Open Business Models. *MIT Sloan Management Review*, Vol. 48, No.2, pp. 22–28.

Collection 42 - University of Oslo Library [WWW Document], 2013. URL http://www.ub.uio.no/english/about/organisation/ureal/ureal/collections/42/index.html.

Cropley, A., 2006. In Praise of Convergent Thinking, *Creative Research Journal*, Vol. 18, No. 3, pp. 391-404.

Culén, A.L., 2015. HCI Education: Innovation, Creativity and Design Thinking. *Proceedings of the Eighth International Conference on Advances in Computer-Human Interactions.* Barcelona, Spain, pp. 125–130.

Culén, A.L., Følstad, A., 2014. Innovation in HCI: What Can We Learn from Design Thinking? *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. Helsinki, Finland, pp. 849–852.

Culén, A.L., Gasparini, A., 2014. Find a Book! Unpacking Customer Journeys at Academic Library, *Proceedings of the Seventh International Conference on Advances in Computer-Human Interactions.* Barcelona, Spain, pp. 89–95.

Culén, A.L., Kriger, M., 2014. Creating Competitive Advantage in IT-intensive Organizations: A Design Thinking Perspective. *HCI Business*, pp. 492–503.

Culén, A.L., et al., 2014. Design, Creativity and Human Computer Interaction Design Education. *Journal on Advances in Life Science,* Vol. 6, No. 3&4, pp. 97-106.

Fallman, D., 2003. Design-oriented human-computer interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Florida, USA, pp. 225–232.

Forlizzi, J., et al., 2008. Crafting a place for interaction design research in HCI. *Design Issues,* Vol. 24, No. 3, pp. 19–29.

Kolko, J., Austin Center for Design, 2012. *Wicked problems problems worth solving.* Ac4d, Austin, Texas.

Koskinen, I., et al., 2011. *Design Research Through Practice: From the Lab, Field, and Showroom* (1 edition). Waltham, MA: Morgan Kaufmann.

Krippendorff, K., 2005. *The Semantic Turn: A New Foundation for Design.* CRC Press, Boca Raton.

Lazar, D.J., et al., 2010. *Research Methods in Human-Computer Interaction.* John Wiley & Sons, NJ, USA.

Leavy, B., 2010. Design thinking – a new mental model of value innovation. *Strategy Leadership,* Vol. 38, No. 3, pp. 5–14.

Lockwood, T., 2009. *Design thinking: integrating innovation, customer experience and brand value*. Allworth Press, New York, NY.

Lynema, E., et al., 2012. Virtual Browse: Designing User-Oriented Services for Discovery of Related Resources. *Library Trends,* Vol. 61, No. 1, pp. 218–233.

Martin, R.L., 2009. *Design of business: Why design thinking is the next competitive advantage.* Harvard Business Press.

Mathews, B., 2012. Think like a startup: A white paper to inspire library entrepreneurialism. Working Paper, Virginia Tech University.

Okun, Y., et al., 2013. BookMotion - INF2260 - Høst 2013 - Universitetet i Oslo [WWW Document]. URL http://www.uio.no/studier/emner/matnat/ifi/INF2260/h13/presentations/BookMotion/.

Realfagsbiblioteket - Google Play [WWW Document], 2013, URL https://play.google.com/store/apps/details?id=no.uio.ub.realfagsbiblioteket&hl=en.

Realfagsbiblioteket - Apple App Store [WWW Document], 2013, URL https://itunes.apple.com/no/app/realfagsbiblioteket/id668535531?mt=8.

Reistad, H.B., et al., 2012. Student Project Bookworms [WWW Document]. URL http://www.uio.no/studier/emner/matnat/ifi/INF2260/h12/projects/library-projects/Bookworms/.

Runco, M.A., Acar, S., 2012. Divergent Thinking as an Indicator of Creative Potential. *Creativity Research Journal,* Vol. 24, No.1, pp. 66–75.

Rundblad, K., 2011. The Academic Library Startup: UX, Innovation and Social Technologies. AJCU Libr. Deans Dir. Conf. Online: http://digitalcommons.lmu.edu/ajcu/2011-conference-program/presentations/4/

Schmidt, A., Etches, A., 2012. *User Experience (UX) Design for Libraries.* Vol. 18. American Library Association.

Seidel, V.P., Fixson, S.K., 2013. Adopting Design Thinking in Novice Multidisciplinary Teams: The Application and Limits of Design Methods and Reflexive Practices. *Journal of Product Innovation Management,* Vol. 30, No. S1, pp. 19–33.

Sharp, H., et al., 2007. *Interaction Design: Beyond Human-Computer Interaction,* 2nd ed. John Wiley & Sons, NJ, USA.

Shelburne, W. A. 2009. E-Book Usage in an Academic Library: User Attitudes and Behaviors. *Library Collections, Acquisitions, and Technical Services,* Vol. 33, No. 2-3, pp. 59–72.

Von Hippel, E., 2009. Democratizing innovation: the evolving phenomenon of user innovation. *International Journal of Innovation Science,* Vol 1, No.1, pp. 29–40.

Wagner, T., Compton, R.A., 2012. *Creating innovators: the making of young people who will change the world.* Scribner, New York.

Zimmerman, J., Forlizzi, J., 2014. *Research Through Design in HCI*, in: Olson, J.S., Kellogg, W.A. (Eds.), Ways of Knowing in HCI. Springer New York, pp. 167–189.

A VIRTUAL KEYBOARD INTERFACE FOR HEAD-MOUNTED DISPLAY DEVICES

Yeonho Kim, Myungchul Sung and Daijin Kim
Pohang University of Science and Technology, South Korea, 77 Cheongam-Ro, Pohang, Gyeongbuk, Korea

ABSTRACT

We implemented a virtual keyboard interface using depth camera and projector. The primary function of this interface is to recognize the typed keys on the virtual keyboard. To make this function possible, we implemented several functions: (1) Generation of virtual keyboard from the projector, (2) Capture of depth image, (3) Detection of hands and extraction of fingertips, (4) Recognition of clicking gesture, and (5) Localization of typed key positions. This interface is useful to generate keyboard commands in any places when the user is walking around.

KEYWORDS

Augmented Reality, Virtual Keyboard, Camera-Projector Calibration, Finger-tip Detection, Gesture Recognition

1. INTRODUCTION

The next generation of input interface for head-mounted display (HMD) requires several features: mobility (convenient portability), natural wearability, immersive reality (user experiences), and user-friendly operability. Continuous improvement of computing device in size and weight makes the mobility and wearability achievable. However, the others require the elaborate design.

Since the first invention of The Sword of Damocles (Sutherland, 1968) by Ivan Sutherland, the lack of input device for HMD has become a significant problem. For the decades, several researchers tried to handle the problem in different ways. In the beginning, a set of sensors is attached to capture the user's indirect command such as head position and orientation (Sutherland, 1968), gaze directions (Park, et al., 2008). However, these kinds of command are just used for a particular purpose (i.e. navigating the user's view) designed by developers. After general purpose sensors such as camera and markers became small and light, some researchers try to capture the user's gesture (Davis & Vaks, 2001) (Freeman & Roth, 1995). However, the commands are complicated and non-intuitive, yet. In the recent work, more direct command are used such as speech recognition (Google, 2014), compatible devices (Olson, et al., 2011), and small-sized buttons (Kato & Billinghurst, 1999), but their expression range was limited.

For immersive reality, the desired system needs to understand the operating environment. This understanding has been studied in terms of augmented reality in computer vision for several years. A variety of approaches are proposed to cope with the problem: fixed projector and camera (Wilson & Benko, 2010), moving projector-camera combination (Raskar, et al., 2004), projecting augmented image onto a predefined object (McFarlane & Wilder, 2009). Most of the attempts requires the elaborate calibration in the environment. Otherwise, a specialized sensor are installed to the interactive objects such as target button or fingers.

Understanding of user's behavior using user-friendly instruments such as hands and arms has been a research topic in computer vision. At the beginning of this field, researchers try to analyze user experience using color information. They have used the human skin model (Von Hardenberg & Bérard, 2001) to detect hands and fingers. Since the skin color is different each other according to individuals, the next attempt was using certain colored gloves (Wang & Popović, 2009) and specialized markers. However, these trials are harmed by low accuracy and inconvenient additional equips. During recent years, development of commercial depth camera (PrimeSense LtD., 2010) makes it feasible to cope with localizing hand and finger positions.

ISBN: 978-989-8533-38-8 © 2015

In this paper, we introduce an easy-to-use virtual keyboard interface for HMD devices. Our proposed system consists of object segmentation, the augmented keyboard projection, finger-tips detection and click gesture recognition. At first, we segment the individual objects from the raw depth image and extract the target plane and foreground objects. The target plane and objects are analyzed to augment a virtual keyboard and localize the finger-tips, respectively. Then we track the position of finger-tips to recognize click gestures and localize positions of the gesture. If the position locates in a certain region on the virtual keyboard, we generate a keyboard input command.

Our proof-of-concept device facilitates the mobility and wearability using a small-sized depth camera and projector. Furthermore, we construct a hand-made helmet using a commercial 3D printer to mount the camera and projector rigidly. To enable interaction with projected the virtual keyboard, we need to calibrate the projector and camera once. The proposed algorithm operate on a hand-held mini PC and its response time is about 60 fps. Our experimental results show the proposed interface allows us to interact with any plane such as a wall and desk, and generates continuous keyboard commands without additional equips such as gloves or markers.

The subsequent sections are structured as follows. Section 2 shows projector-camera calibration that is required to perform once before the algorithm. We describe the details of object segmentation and hand/finger localization in section 3 and show their experimental results in section 4. Finally, section 5 presents the conclusion.

2. SYSTEM CONFIGURATIONS

In this section, we describe the details of the hardware configuration of virtual keyboard interface and the camera-projector calibration.

2.1 Hardware Configuration

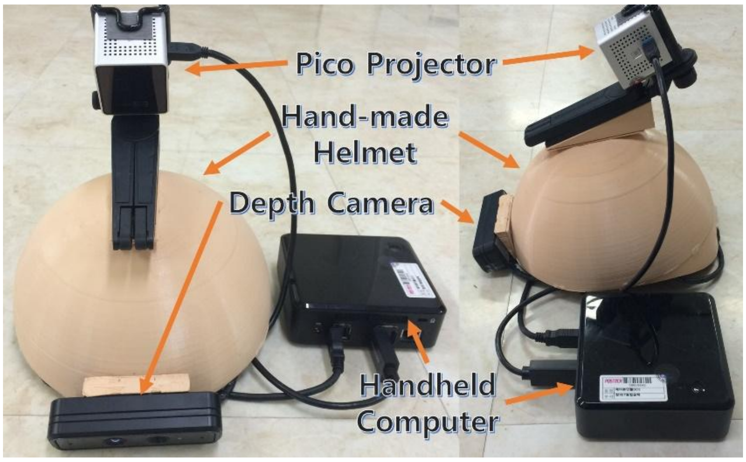


Figure 1. Hardware configurations of our proposed system.

The proposed system, Virtual Keyboard Interface, consists of four components as shown in Figure 1. The first one is a short-range, small-sized depth camera, DepthSense, which captures a depth map in every 30ms. The camera captures objects within 1m from the sensor. The camera size of $10.5 \times 3 \times 2.3$cm suits for mobility and convenience of our hardware implementation. The sensor measures the round-trip time it takes for infrared light emitted from the camera and produces the depth map of 320×240 pixels. These pixels represent three-dimensional Cartesian coordinates (X, Y, Z), with 1.4cm of depth resolution. The origin of the coordinate system is located in the center of the IR sensor. The second component is an INNOIO SK Smart Beam handheld projector; its size is $4.5 \times 4.5 \times 4.5$ cm, and weight is 129 g. Since this small-sized, battery-powered projector runs about 2 hours, the projector suits for our prototyping system. Moreover, since the projection image stretches about 31 inches (i.e. 61×46cm) at 1m from the camera, our system provides a full-sized virtual keyboard. We rigidly mount both the camera and projector on a hand-made helmet, the third component, built by utilizing a commercial CubeX 3D printer. The future expectation of usage in HMD made the choice of the hemisphere mounting frame. Our choice make users can locate the virtual keyboard on their demand and be free from occlusion of the camera and projector. Though, an unrestricted movement of the head causes the projected keyboard jitter, each of the buttons has an enough size to overcome the problem. Finally, both the camera and project are wired to a handheld mini PC.

2.2 Camera-Projector Calibration

To interact with the projected virtual keyboard in the immersive reality, it is necessary to calibrate the camera and projector, which are treated as one base camera and its corresponding inverse camera. Since the typical process of camera calibration requires more than two cameras, not a projector, we regard the projector as a particular camera. After that, the calibration process requires calculating the intrinsic parameters of the sensors and the extrinsic parameters among them. The intrinsic parameters, which consists of focal length, principal point, and skew coefficient, are formulated as 3×3 matrices of a set of constant values unless the sensors' configuration are changed. While, the extrinsic parameters define the relative position and orientation of the sensors. Therefore, the extrinsic parameters are easily changed according the sensor's configurations. To calculate the extrinsic parameters, we need to collect some corresponding calibration points for both the camera and the projector. Since it is more straightforward for the depth camera to find a number of real-world points than the projector, we fix the calibration points in the projector space and try to find the corresponding points in the camera space.

The projector emits a chess-board image in an arbitrary plane such as a wall or desk for finding the corresponding calibration points. Then, the projected chess-board is captured by the camera. The corner detection algorithm reveals the calibration points at the chess-board image in the projector and the corresponding points at the chess-board image in the camera. Once the correspondence is established, the camera points are transformed to 3D coordinates in the camera space. Note that we extract the calibration points in one arbitrary plane so far. To cover possible plane spaces, we need to repeat the collecting sub-process into other planes as possible. In our experience, it is sufficient for collecting ten sets of points to calculate the accurate correspondence. (3dsense, 2013).

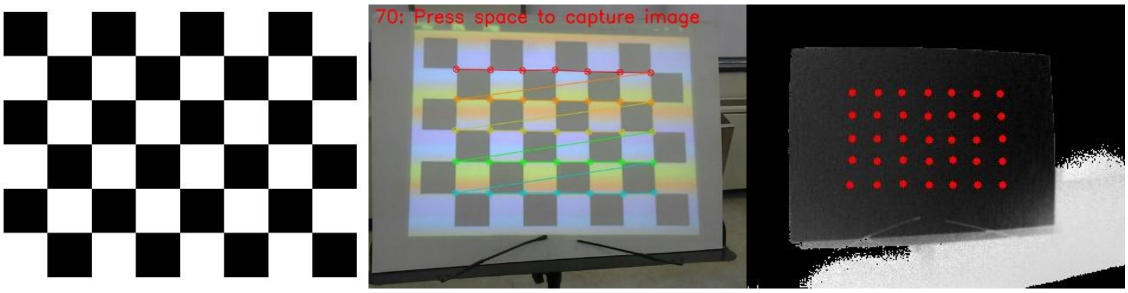


Figure 2. The example images of camera-projector calibration: (1) chess-board image in the projector, (2) chess-board image in the color sensor of the camera, (3) the transformed 3D points in the camera.

To calculate the extrinsic parameters of the sensors, we use the pinhole camera model, which is represented in the Homogeneous coordinates:

$$z_o X_p = A[R|T]X_c, \tag{1}$$

where we denote a projector point as $X_p = (x_p, y_p)^T$, its corresponding point in the camera space as $X_c = (x_c, y_c, z_c)^T$, the projector intrinsic matrix as A, the rotation matrix of projector as R, and the translation vector as T. Since we assume the camera space as world space, note that, we put the camera point same as the world point: $X_c = X_w = (x_w, y_w, z_w)^T$. By expanding the above equation, we get the following equation:

$$z_o \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 r_{11} & c_1 r_{12} & c_1 r_{13} & c_1 t_1 \\ c_2 r_{21} & c_2 r_{22} & c_2 r_{23} & c_2 t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \tag{2}$$

ISBN: 978-989-8533-38-8 © 2015

To make the number of unknown variables same as the degree of freedom, we divide the elements of the matrix by t_3 and replace with coefficients, $q_1 \dots q_{11}$:

$$z_o \begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \\ q_5 & q_6 & q_7 & q_8 \\ q_9 & q_{10} & q_{11} & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}. \qquad (3)$$

Now the above equation is expanded again, and we get two equations,

$$\begin{aligned} x_p &= q_1 x_c + q_2 y_c + q_3 z_c + q_4 - q_9 x_c x_p - q_{10} y_c x_p - q_{11} z_c x_p, \\ y_p &= q_5 x_c + q_6 y_c + q_7 z_c + q_8 - q_9 x_c y_p - q_{10} y_c y_p - q_{11} z_c y_p \end{aligned} \qquad (4)$$

For the every corresponding point, we make a matrix notation using the above equations:

$$\begin{bmatrix} x_k^1 & y_k^1 & z_k^1 & 1 & 0 & 0 & 0 & 0 & x_k^1 x_p^1 & y_k^1 x_p^1 & z_k^1 x_p^1 \\ 0 & 0 & 0 & 0 & x_k^1 & y_k^1 & z_k^1 & 1 & x_k^1 x_p^1 & y_k^1 x_p^1 & z_k^1 x_p^1 \\ x_k^2 & y_k^2 & z_k^2 & 1 & 0 & 0 & 0 & 0 & x_k^2 x_p^2 & y_k^2 x_p^2 & z_k^2 x_p^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & x_k^n & y_k^n & z_k^n & 1 & x_k^n x_p^n & y_k^n x_p^n & z_k^n x_p^n \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_{11} \end{bmatrix} = \begin{bmatrix} x_p^1 \\ y_p^1 \\ x_p^2 \\ \vdots \\ y_p^n \end{bmatrix} \qquad (5)$$

To solve this system, we use QR decomposition, which minimizes least square error. Now we can transform the points in the camera space to the points in the projector space with these eleven coefficients:

$$x_p = \frac{q_1 x_c + q_2 y_c + q_3 z_c + q_4}{q_9 x_c + q_{10} y_c + q_{11} z_c + 1}, \quad y_p = \frac{q_5 x_c + q_6 y_c + q_7 z_c + q_8}{q_9 x_c + q_{10} y_c + q_{11} z_c + 1}. \qquad (6)$$

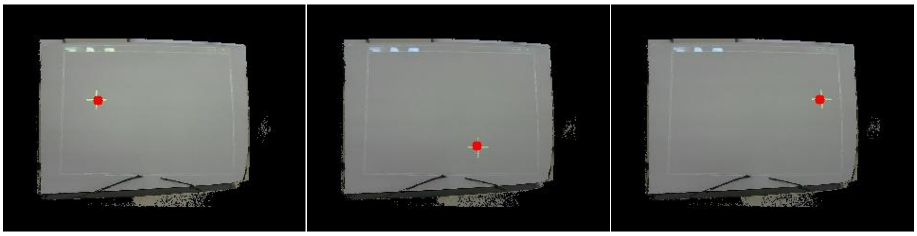


Figure 3. The validation of camera-projector calibration: red circles and overlapped green crosses represents points in the camera space and transformed points into the projector space, respectively.

3. PROPOSED METHOD

Figure 4 shows an overview of our proposed system for a virtual keyboard interface, which consists of four sub-processes: object segmentation, finger-tip detection, gesture recognition, and keyboard command generation. The subsequent sub-sections present the details of the proposed algorithm.

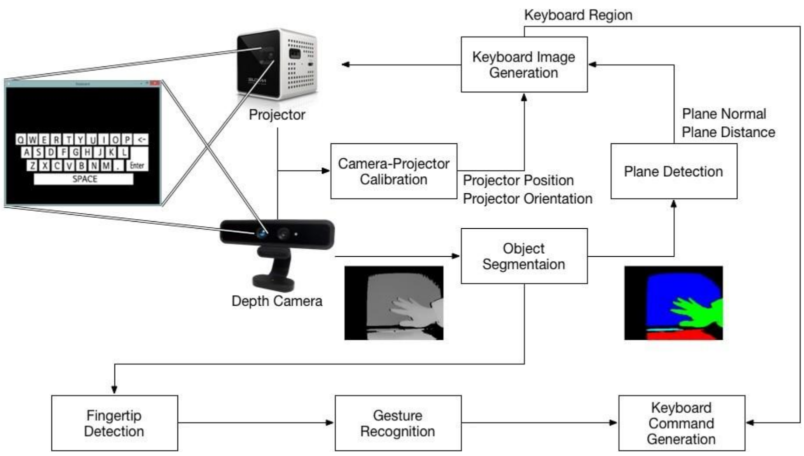


Figure 4. Overall proposed process for a virtual keyboard system.

3.1 Object Segmentation

To identify a target plane and hands in depth images delivered from the depth camera, we segment each object using connected component labeling(CCL). The typical CCL algorithm is one of typical segmentation technique and gives a unique label to each connected component according to its connectivity criteria. The algorithm determines the pixel connectivity as whether there are binary pixels in neighboring pixels or not. To apply the CCL algorithm into the depth image, we modify the connectivity criteria to consider the difference in depth values of neighboring pixels. For example, when one's hand hovers above a plane, the boundary pixels of the hand have similar depth values with pixels in the hand but have smaller values than the plane. The 3D CCL, however, sometimes produces unexpected large-sized objects, when some accidental depth data is located in the middle of two objects. The noisy data makes the two objects be collected, unfortunately. To eliminate the problem, we adapt a novel noise reduction method. After applying 5×5 pixeled MIN and MAX filters, simultaneously, we monitor the difference between the two results. We see the insignificant difference in a smooth region such as arm and plane while see the distinct difference in the boundary region.

We classify the segmented objects into three types: non-interest objects, plane object, and arm objects. Since the small-sized objects are negligible, we filter out the non-interest objects, which are smaller than a hand. The second type of the segmented objects is a plane object. The farthest object, which has significant depth values comparing to other objects, is considered as the plane object. At the last, we select at most two objects that have small depth values among the remaining objects, as arm objects. We pass the arm objects to the finger-tips detection as an input.

The plane object, one of the key objects, has useful information that is translated to generate the augmented virtual keyboard. Since our system allows the users locate a virtual keyboard on their demand, we have no idea that where the target plane faces or how far the plane is. To make a proper virtual keyboard, we utilize the three-dimensional information of the plane object. The Principal Components Analysis (PCA) algorithm extracts the axis information of the target data. The first two components represent the directions where the plane spans while the third component represents the normal direction of the plane. Also, the mean vector of the data is used as the center position of the plane. We update a Homogeneous matrix, which rotates and translates an initial keyboard image into a target keyboard image, whenever the normal vector or the center position is changed.

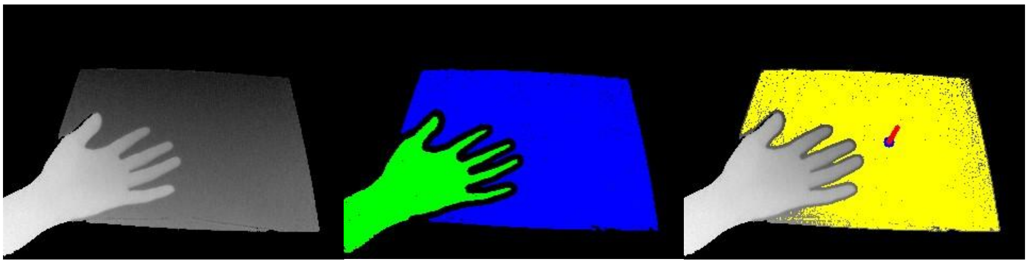


Figure 5. Objects Segmentation: (1) the raw depth image, (2) the segmented objects, (3) the plane object (yellow) and its normal vector (red line).

3.2 Finger-tips Detection

Locating finger-tips in the segmented arm objects has four steps; edge detection, feature extraction, Geodesic extreme calculation, and finger-tips detection. First, we take a depth map of one segmented arm object and detect the edge. We then transform the depth map into a distance map, which presents the distance from a closed edge and compute a probability of our feature, Ridge data (Jalal & Kim, 2014), using various sized sliding filters. The filter size is determined by the distance value of the distance map:

$$R(I) = \left\{ X_c \in I \,\middle|\, \frac{\sum_{i=1}^{N} D_T(X_i)}{N \cdot D_T(X_c)} < \delta_R \right\}, \tag{7}$$

where N is the number of reference points surrounding the center point X_c, $D_T(\cdot)$ is the distance value at the point, and δ_R is a certain threshold of Ridge probability. The Ridge data (white dots in Figure 6) acts as a set of finger-tip candidates. Third, we compute the Geodesic distance (Baak, et al., 2013) of the Ridge data to avoid locating finger-tips in the object's boundary. The Dijkstra's algorithm computes shortest path between two positions in greedy manner. In order to extract a number of Geodesic extremes, however, the algorithm has redundant computations. By simply adding a new edge after each iteration, we can obviate the redundancy problem. At last, we consider the closed Ridge data to each of Geodesic extremes as finger-tip.

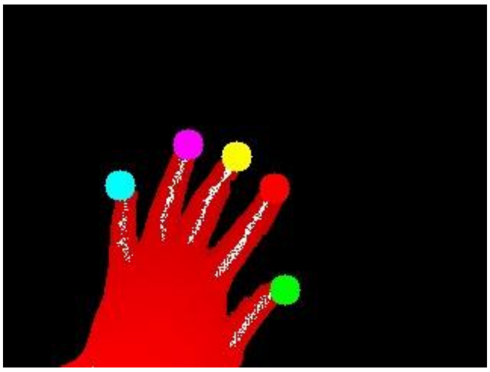


Figure 6. The representation of Ridge data(white dots) and detected finger-tips(colored circles).

3.3 Gesture Recognition and Keyboard Command Generation

To recognize the click gesture, we monitor the distances of the tracked finger-tips from the target plane. We compute the distance derivative in a particular period. If the zero-crossing of the derivative occurs, we know a click gesture. Every click gesture, however, makes the finger-tips move, not only the desired finger-tip but the neighbor finger-tips. Therefore, we monitor the change of the distances with fixed size of sliding window and consider the biggest movement among the monitored finger-tips shows the desired gesture.

Figure 7 exemplify the situation and the finger-tip, f2, make a biggest movement from 14 to 19.

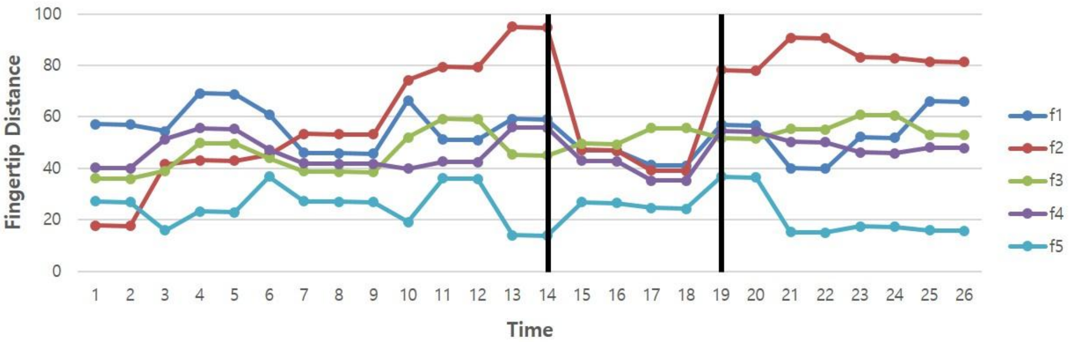


Figure 7. The representation of the changes of finger-tip distance over time: each of colored lines shows the distance of individual finger-tip.

During the monitoring the distance of the finger-tips is enough to recognize click gestures, we apply additional conditions to differentiate just a moving hand and actual click gestures. An ordinary hand movement like hovering and locating hands in a certain region could generate click gestures as consider just the distance of the tracked finger-tips. The moving direction of finger-tips is the most significant different between ordinary movements and click gestures. To understand this difference, we regard the hand movement toward the spanning direction of the target plane as non-gestures.

The final step of our proposed algorithm is localizing the click gesture. Whenever the click gestures are recognized in a certain distance from the target plane, we transform the gesture positions in 3D coordinates into the projector positions in 2D coordinates. This transformation is calculated using the eleven coefficient, which is described in section 2. If the transformation locates the gesture in a certain region of the projected virtual keyboard, we generate a corresponding key event.

4. EXPERIMENTAL RESULTS

We implement our proof-of-concept interface with C++ and run our experiment on a commercial handheld mini PC, Intel NUC Kit, with a 1.8 GHz CPU. We use Our experiments are conducted at the indoor environment and show the proposed algorithm runs at about 60 fps. An analysis of the algorithm shows the major bottlenecks of our algorithm is 3D CCL and Dijkstra algorithm.

For the first experiment, we integrate our proof-of-concept interface with Google Glass. Two individual systems are connected via a wireless network, and we demonstrate the feasibility of future HMD usage. As shown in Figure 8, our proposed system fits an existing HMD device and its usage covers from a simple messaging to writing a complex document.

Figure 8. The example images of our demonstration integrated with Google Glass

The next experiment shows the results of virtual keyboards, which suit to the target plane. We see misaligned keyboards and those window frames at both images in Figure 9. The result of this experiment means our virtual keyboards are fit for the target plane, but the window frames are not.

Figure 9. The experimental results of the augmented virtual keyboard.

5. CONCLUSION

In the paper, we introduced our virtual keyboard interface for HMD. Our accurate camera-projector calibration has unlimited usage toward augmented reality area. Moreover, we believe that a new representation of finger-tips will help understanding human behaviors in many research area. A relatively narrow view of both the camera and projector make user interact in a restricted region, the existing moving camera compensation technique will reduce the difficulty with help of wide field-of-view of camera and projector in the future. While we have a set of heuristics, we will gradually remove them by intensive user tests and user experience learning.

ACKNOWLEDGEMENT

This research was partially supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the "ICT Consilience Creative Program" (IITP-2015-R0346-15-1007) supervised by the IITP(Institute for Information & communications Technology Promotion) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2010-0019523)

REFERENCES

3dsense, 2013. [Online] Available at: http://blog.3dsense.org/programming/kinect-projector-calibration-human-mapping-2/

Baak, A. et al., 2013. A data-driven approach for real-time full body pose reconstruction from a depth camera. s.l., Consumer Depth Cameras for Computer Vision.

Davis, J. W. & Vaks, S., 2001. A perceptual user interface for recognizing head gesture acknowledgements. s.l., the 2001 workshop on Perceptive user interfaces.

Freeman, W. T. & Roth, M., 1995. Orientation histograms for hand gesture recognition. s.l., International workshop on automatic face and gesture recognition.

Google, 2014. [Online] Available at: https://www.google.com/glass/start/

Jalal, A. & Kim, Y., 2014. Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data. s.l., Advanced Video and Signal Based Surveillance .

Kato, H. & Billinghurst, M., 1999. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. s.l., Augmented Reality.

McFarlane, D. C. & Wilder, S. M., 2009. Interactive dirt: increasing mobile work performance with a wearable projector-camera system. s.l., the 11th international conference on Ubiquitous computing.

Olson, J. L., Krum, D. M., Suma, E. A. & Bolas, M., 2011. A design for a smartphone-based head mounted display. s.l., In Virtual Reality Conference .

Park, H. M., Lee, S. H. & Choi, J. S., 2008. Wearable augmented reality system using gaze interaction. s.l., the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality.

PrimeSense LtD., 2010. [Online] Available at: http://www.primesense.com

Raskar, R. et al., 2004. RFIG lamps: interacting with a self-describing world via photosensing wireless tags and projectors. ACM Transactions on Graphics, 23(3), pp. 406-415.

Sutherland, I. E., 1968. A head-mounted three dimensional display. s.l., fall joint computer conference.

Von Hardenberg, C. & Bérard, F., 2001. Bare-hand human-computer interaction. s.l., the 2001 workshop on Perceptive user interfaces.

Wang, R. Y. & Popović, J., 2009. Real-time hand-tracking with a color glove. ACM Transactions on Graphics, 28(3), p. 63.

Wilson, A. & Benko, H., 2010. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. s.l., UIST.

BRAIN COMPUTER INTERFACES FOR MOBILE APPS: STATE-OF-THE-ART & FUTURE DIRECTIONS

Sumit Soman[1], Siddharth Srivastava[1], Saurabh Srivastava[2] and Nitendra Rajput[3]
[1]*Department of Electrical Engineering, Indian Institute of Technology, Delhi, India*
[2]*Xerox Research Centre, Bangalore, India*
[3]*IBM India Research Laboratory, New Delhi, India*

ABSTRACT

In recent times, there have been significant advancements in utilizing the sensing capabilities of mobile devices for developing applications. The primary objective has been to enhance the way a user interacts with the application by making it effortless and convenient. This paper explores the capabilities of using Brain Computer Interfaces (BCI), an evolving subset of Human Computer Interaction (HCI) paradigms, to control mobile devices. We present a comprehensive survey of the state-of-the-art in this area, discussing the challenges and limitations in using BCI for mobile applications. Further we propose possible modalities that in future can benefit with BCI applications. This paper consolidates research directions being pursued in this domain, and draws conclusions on feasibility and benefits of using BCI systems effectively augmented to the mobile application development domain.

KEYWORDS

Human Computer Interaction (HCI), Brain Computer Interface (BCI), Electroencephalogram (EEG), Mobile Interfaces, Assistive Technology

1. INTRODUCTION

Advancements in mobile hardware and sensing capabilities have drastically altered the scope of using mobile devices as well as the procedure in which the user interacts with them. Multiple communication modalities, effective HCI paradigms and a versatile communication network have paved the way for building mobile applications which interact with the user in a more ”natural” manner. The next development appears to be the use of consciously modulated parameters of the human body to communicate with mobile apps, which would make user interaction seamless and intuitive. As one such use case, BCI as a means for HCI on mobile devices is beneficial to enable communication for patients with locomotor disabilities. BCI presents a viable alternative, as conventional communication channels are not available for interacting with mobile platforms. Drawing motivation from such scenarios, we seek to explore the state-of-the-art in the use of BCI as a modality for providing user input to mobile applications.

There are several challenges in building mobile applications based on BCI systems. A key challenge is the choice of control paradigm, or precisely, what actions should the user perform which would be detected by the BCI system to execute actions on the mobile application. This is a difficult choice due to the variations in reproducibility of the control intents across users, which directly affects the classification accuracy of the BCI system. Another important challenge is the platform for processing of EEG signals. The EEG signals from acquisition devices are represented as floating point matrices, which have high dimensionality. To provide a representative estimate, EEG data acquired from 32 channels (recording locations) for 5 seconds at a sampling rate of 250 Hz would be represented by a matrix of size $[32 \times 1250]$. Operations on large matrices require high computational cost. However, the computational power available on mobile platforms is limited, hence the signal-processing pipeline needs to be optimized to suit the platform architecture. The challenges in the use of BCI for gaming applications have been discussed in the work by Nijholt et al. [40, 39, 38]. Despite these, BCIs have also been used in developing assistive technology [8], mobile robot control [3, 37] and many other applications [36, 23, 19].

The objective of this paper is to present a survey of the recent work done in the field of applications developed for mobile platforms based on BCI. To the best of our knowledge, this would be a pioneering effort in this direction, which will enable consolidation of research being pursued and draw conclusions on feasible future directions. The rest of the paper is organized as follows. Section 2 discusses the details of the survey methodology used in this paper, followed by the characteristics required in an EEG acquisition system feasible for development of mobile applications in Section 3. This is followed by a discussion on the modalities employed in developing BCI based mobile applications along with a few BCI based mobile applications in Section 4. The conclusions and future directions are presented in Section 5.

2. SURVEY METHODOLOGY

The content presented in this paper is based on research papers and articles that are indexed and available on Google Scholar and Scopus. Papers with index terms using combinations of "Mobile" with "BCI", "Assistive Technology", "EEG", "Brain Machine Interface (BMI)" and "HCI" have been discussed in the review. The focus has been to lay more emphasis on papers, which have appeared in the past 3 years, though this has not been strictly adhered to in cases where seminal work has been indicated.

3. BUILDING BCI SYSTEMS FOR MOBILE: DEVICES & MODALITIES

Developing BCI systems for practical mobile applications involves several challenges owing to portability and degree of involvement of the user in triggering control intents. Conventional EEG acquisition devices need to be made portable and ergonomic, while control paradigms should need minimal effort and not cause fatigue. This section presents details on characterizing an "ideal" acquisition device, and discusses some of the common paradigms used in such applications.

3.1 EEG Acquisition Systems for Mobile

One of the primary challenges faced while using BCI for mobile platforms has been the choice of the EEG Acquisition System. Conventional EEG acquisition systems suffer with the following constraints from the perspective of usability and flexibility over mobile devices:

1. They are large and non-portable.
2. They have large number of electrodes (EEG acquisition sensors).
3. Many interconnecting wires to connect the amplifier with the electrodes.
4. Requirement of a trained professional for connecting electrodes as per the 10-20 system [22].
5. Usage requires application of a conductive gel [50] or saline.

Such systems are more suited for EEG-based investigations in medical facilities. Due to these constraints, they cannot be directly used for mobile-based EEG applications. However, the benefit lies in the higher resolution of the EEG signals obtained from them in terms of number of channels, lower noise levels, higher sampling rates and the ability to configure the channel montage. This in turn aids the machine learning pipeline to achieve higher classification accuracies. Hence, an ideal mobile EEG acquisitions device would have a compact form factor, be easy to wear and setup, lightweight and enable real-time processing of EEG signals. It would have a convenient and wireless connection interface, preferably over Bluetooth or Wi-Fi. High-quality signals should be obtained, in terms of frequency resolution and low-noise.

In view of the challenges identified above, newer EEG systems have been developed. Gargiulo et al. [14] demonstrate an EEG system, called Penso, with a compact form factor and Bluetooth connectivity. The system has shown high correlation coefficient of 0.83 on average with clinical grade EEG in terms of alpha [27] and mu [42] rhythm detection for 8 subjects. It has dry electrodes made with conducting rubber, which enable it to be setup easily. The device has a measurement bandwidth of 0.4-40 Hz, and could be setup with an average setup time of 10 seconds per electrode. This is lower than the time required for a system with wet electrodes, which would be around 2-3 minutes per electrode. This reduction in setup time is significant.

Further, Chi et al. [7] have developed a dry and contact-less EEG system for use in mobile applications. These electrodes do not require direct contact with the scalp and are based on capacitive coupling with the scalp through the hair. They also have active buffering capability; high impedance and effective shielding against noise, which provides improved quality EEG signals. The developed system has been shown to achieve an Information Transfer Rate (ITR) of 19 bits/minute with 100% accuracy for the SSVEP modality. Another popular mobile EEG headset has been the Emotiv EPOC headset [48]. It is a 14 channel wireless headset with saline-based electrodes. The headset has been ergonomically designed and is easy to wear and setup as it has a wireless interface. It has been used in a variety of gaming applications [53, 44] and assistive technologies [30, 17]. The headset also has a more compact version called the "Emotiv Insight" [1]. Other portable systems include the Neurosky and G.MOBIlab [16]. A review of the transition from laboratory to mobile BCI has been discussed in the work by Edlinger [13], and requirements of a portable system have been enumerated by Debener et al. [12]. A review of EEG systems used for BCI can be found in the work by Lin et al. [32].

3.2 BCI Control Modalities for Mobile Applications

This section aims to discuss the modalities available and in use for developing mobile applications based on BCI. We consider gestures and eye movements, SSVEPs and P300 as primary control paradigms.

3.2.1 Gestures and Eye Movements

The ability to detect facial gestures, eye-blinks and eye movements from the EEG signals makes it a convenient paradigm to enable control in mobile applications. This modality encompasses a variety of control triggers, including eye blinks, lateral eye movements, facial gestures such as wink, twitch, jaw clench and smirk. The benefit of using these lies in the fact that a variety of intents are available, each of which translates into direct actions. Hence, it offers higher degree of freedom in terms of multi-class BCI. These can also be integrated with other modalities like eye gaze tracking [49].

3.2.2 SSEVP

Steady State Visually Evoked Potentials (SSVEPs) have been used to develop several BCI systems. The key concept is that the user's brain responds to stimulus presented at specific frequencies, which may be realized by flashing the commands to the user at different frequencies. When the user focuses attention on the object of interest, the brain will generate signals at that frequency. The advantage is that this is less susceptible to artifacts, and reproducible across subjects [2, 54]. The downside is that users may find the stimulus overwhelming, as it needs to be flashed at different frequencies. For instance, [60] indicates that the use of SSVEPs may induce seizures or impair vision. Recent works on adapting SSVEPs include the development of stimulus presentation paradigms for mobile applications [56] and testing the versatility of SSVEPs while users are engaged in other physical activities [34]. In [35], the authors present a feasibility study of using SSVEPs when the subject is engaged in walking. This shows promise in the use of this paradigm. Examples of SSVEP based BCI applications for mobile have been discussed in Section 5.

3.2.3 P300

The P300 signal is another well-known BCI control intent. It represents EEG signals from the parietal lobe, which is known to occur 300 milliseconds after an "odd-ball" event. This has been used in the P300 speller, where the user is shown a grid of characters in which the rows and columns are flashed in a predetermined sequence. The user is asked to focus on the letter of interest and count the number of times it is flashed. This is termed as an "odd-ball" event, which leads to the generation of a P300 signal. It can be detected by EEG signal processing to identify the user intent. In this manner, users are able to spell words using the P300 speller. Efforts have been made to study the versatility of using P300 [43], in the context of mobile applications as an extension of the P300 speller with predictive text [21], and in combination with auditory stimulus by Vos et al. [10, 11] Their study for a 3-class auditory oddball task on 20 subjects gave average classification accuracies of 71% for sedentary and 64% for walking states. The task used a standard tone of 900 Hz and two deviant tones of 600 and 1200 Hz with a mean inter-stimulus interval of 1 second and jitter between 0-375ms. The data analysis involved re-referencing, Extended Infomax ICA [28], and classification using Stepwise LDA [25].

Notable conclusions and inferences were presented by the authors in [10,11]. They opined that practical BCI systems should be robust, reliable and easy to setup outside the controlled laboratory environment. It has been inferred that lower accuracies of BCI systems while subjects are performing physical actions can be attributed to the fact that the brain is involved with other tasks. As P300 amplitudes are a measure of the attention levels, these decrease in such cases. An important proposition to resolve this is multi-modal BCI, which incorporates feedback from other paradigms such as eye tracking, head movement (gyroscope) and video recording of visual scene. However, the practical realization of any multi-modal system would also demand computationally intensive processing. Hence, there is a need to develop such a processing pipeline, which is an active area of research. For instance, minimal complexity classifiers have been investigated in [26, 24].

4. BCI FOR MOBILE APPS: STATE OF THE ART

The SmartPhone Brain Scanner (SBS) app developed by Stopczynski et al. [45, 46, 47] provides a hand-held brain scanner on mobile using the Emotiv headset for real-time image reconstruction of brain activity. This application serves as an assistive tool for EEG diagnosis in remote areas. The open source implementation allows extension of the framework to other BCI-based mobile applications. It provides multiple feature extraction and classification methods, making this a candidate platform for BCI-based mobile application development. As the authors indicate in [45], SBS does not have automatic artifact detection or elimination methods, which are crucial for efficient EEG signal processing. On similar lines, Lin et al. [33] perform remote transmission of brain images over communication networks using client-server architecture. The objective is a mobile visualization tool, with focus on efficient structures for storage, transmission and re-construction of the brain images over the network to remote devices.

We now discuss day-to-day applications based on BCI for mobile platforms. Neurophone, by Campbell et al. [6] is a phone-dialer application based on the P300 paradigm. Users are shown images of contacts flashed randomly and the P300 stimulus is used as a trigger to initiate a call with the contact. Their paper presents two versions of this application, one that is triggered by eye-winks and another that is triggered using P300. The paper concludes that the “wink-triggered” version works robustly even in noisy conditions, whereas the ”thought-triggered” dialer appears “promising but at present less reliable.” The application has been tested on three subjects under sedentary and walking conditions, and reports accuracies in the range of 92-95% for the “wink-triggered” version and lower accuracies for the P300 version. This work is novel as a working application has been developed and evaluated rigorously.

A P300 based BCI Messenger has been developed by Li et al. [29] as a tool for day-to-day communication, and has achieved over 80% accuracy on 6 subjects across sessions. The application consisted a [6 × 6] grid with Chinese strokes to be used as a virtual keyboard. The training phase involved the subjects viewing specific targets on the screen while their EEG was being recorded, and a subsequent phase where visual feedback was provided to this task. The subjects undertook 3 sessions of training with around 20 runs per session. The paper reports an average accuracy of 69.7%, and suggests possible improvement by use of larger-sized visual stimulus, use of channel selection to reduce EEG dimensionality and use of alternate classifiers. The *CharStreamer* paradigm proposed by Höhne et al. [20] aims to incorporate auditory and visual stimulus in order to realize a speller using Event Related Potentials (ERPs).

Several BCI based mobile applications using the SSVEP paradigm are available in literature. Wei et al. [58] develop an SSVEP based user calling application with an LED-based visual simulator panel for eliciting user input. The input is mapped to key-presses on the standard numeric keypad with dedicated keys for call receiving/disconnecting. The user interface consists of a [4 × 4] grid with 10 number keys and 6 function keys for dial, disconnect, answer, fast dial, enter (to complete current operation) and backspace.

Table 1. A comparison of recent works in the area of BCI-based applications for mobile platforms

S.No.	Application	BCI Paradigm	EEG Headset	Mobile Device	Preprocessing &Feature	Classification	Remarks
1	Neurophone [6]	P300	Emotiv	I-Phone	Band Pass Filtering	Multivariate Equi-prior	Contact caller App
2	Smartphone Brain Scanner	-	Emotiv	Multi-platform	Multiple - Fast Fourier	Bayesian Formulations,	Brain activity Imagery App
3	Cell phone BCI [55]	SSVEP	MEMS Sensors	Nokia/ J2ME	FFT, Canonical	Information Transfer Rate	Phone Dialer
4	BCI Messenger [29]	P300	Neuroscan QuickCap	-	Artifact Removal by	Template Matching,	Chinese and English
5	Char Streamer [20]	ERP	EasyCap GmbH	-	Artifact Removal and	Binary Linear Discriminant	Spelling interface
6	BCI Telepresence	Motor Imagery	G.Tec	Robotino	Laplacian spatial filtering,	Statistical Gaussian	Telepresence mobile robot
7	Hex-o-spell [4]	Motor Imagery	Berlin BCI	-	CSP (as in BBCI)	LDA (as in BBCI)	Brain actuated
8	Mental Telephone	SSVEP	Self made	SIM300	CCA	Based on CCA coefficients	Phone dialer app
9	Unlock Project [5]	SSVEP	G.Tec, Emotiv,	Android	Not Specified	Not Specified	Framework for
10	Mobile Device Control [52]	S-EMG	Motion Lab Systems	Android	Band Pass Filtering&	Weighted bandpower	Mobile app to control
11	Mobile Controller [9]	S-EMG	Arm-band for EMG		Not specified	Not specified	Subtle Gesture based device

The system is based on detecting SSVEPs from occipital electrodes, and uses frequency screening to identify feasible frequency bands for specific users. During training, users are presented stimuli at frequencies of 6-20 Hz with a window of 0.5 Hz, and an optimal band is identified using canonical correlation analysis. The developed system has been tested on nine volunteers and reports mean ITR of 33.53 bits/min, which can be attributed to the training process that users undergo prior to online system usage.

Wang et al. [56] design a Phone Dialer app with an average ITR of 26.46 bits/min. A general platform for developing SSVEP based BCI systems with minimal delay has also been developed by Brumberg et al. [5]. It supports multiple EEG acquisition devices and provides a platform agnostic framework. It is evident from the work in this direction that SSVEP based BCI mobile applications have high potential in the future. Ideas for shared brain control in mobile platforms has been explored in the work by Tonin et al. [51]. BCI can also be integrated in a multimodal fashion for interaction with mobile devices [18]. More recently, Lin et al. [31] have demonstrated the use of Micro-Electro Mechanical Systems (MEMS) based EEG sensors to realize wireless and non-invasive prosthesis control. A review of application of BCIs for control of mobile peripherals has been presented by Li et al. [41]. Work has also been done using modalities other than EEG signals, such as the Surface Electromyogram (S-EMG), for gesture based device control [9] and household electronic devices control [52]. A summary is given in Table 1.

5. DISCUSSION & FUTURE DIRECTIONS

This paper presented a survey of BCI based applications developed for mobile platforms. Despite several challenges, BCI research has evolved to many practical applications. This modality offers a natural extension to how humans interact with the external environment, and hence presents a promising avenue.

Figure 1 shows a graph of the number of research articles on mobile-based BCI over the last ten years. It is evident that the trend indicates increasing favor of these research directions, especially over the last six years. Other avenues of interest appear to be developing new paradigms for the mobile devices using EEG signals, for instance the P300 speller [15]. Extensions such as the mental typewriter Hexospell [4] can be developed for mobile platforms, based on probabilistic [59] or multi-modal [57] text entry. SSVEP based applications hold a promising future, owing to the natural adaptability of this paradigm to a menu-based interface. For example, the home menu on any mobile platform with options for contacts, messaging, games, settings, gallery and the like can be implemented as a grid of flashing options, where selection is triggered by SSVEP.

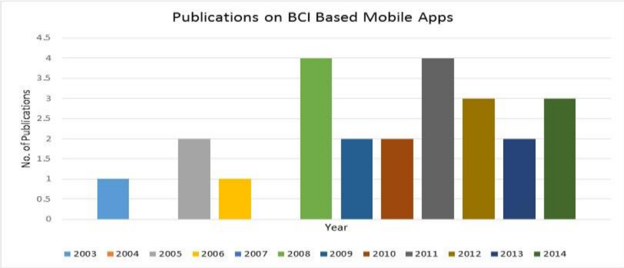


Figure 1. Research papers published over the last ten years in the area of developing mobile applications using BCI.

Mobile devices have several sensors such as accelerometer, gyroscope, compass, barometer, proximity sensors etc., which provide data that can be operated upon by machine learning techniques to augment the user intent. This includes techniques for processing images from camera feed or audio processing from the microphone. The combination of these can be used as an effective measure against the false-positives identified by the BCI modality. User-specific data is also available on mobile devices, which can be harvested to augment BCI-based techniques. Further, data from of mobile usage patterns such as touch, call, applications etc. can be used based on the problem at hand. For instance, the most likely contacts whom the user may call can be determined using call logs and the user can be presented these callers to choose from using the BCI modality (as the first level of choice). The use of BCI for mobile holds promise, as it would be useful for developing both assistive and general entertainment technology for the evolving mobile user. Developments in the fields of computational neurology, signal processing, machine learning coupled with the increasing availability of computational power on the mobile device along with cloud storage and processing paves the way for development of more BCI based mobile applications for the ordinary user.

REFERENCES

[1] (2014) Emotiv insight neuroheadset URL - *http://www.emotiv.com/insight.php*

[2] Allison B., Luth T., Valbuena D., et al, (2010) BCI demographics: How many (and what kinds of) people can use an SSVEP BCI*? IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 18(2):107–116.

[3] Bell CJ., Shenoy P., Chalodhorn R., et al, (2008) Control of a humanoid robot by a noninvasive brain–computer interface in humans .*Journal of neural engineering* 5(2):214.

[4] Blankertz B., Dornhege G., Krauledat M., et al, (2006) The Berlin brain-computer interface presents the novel mental typewriter hex-o-spell.

[5] Brumberg JS., Lorenz SD., Galbraith BV.et al, (2012) The unlock project: A python-based framework for practical brain-computer interface communication app development. *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, IEEE*, pp 2505–2508.

[6] Campbell A, Choudhury T, Hu S, et al, (2010) Neurophone: brain-mobile phone interface using a wireless EEG headset. *Proceedings of the second ACM SIG- COMM workshop on Networking, systems, and applications on mobile handhelds*, ACM, pp 3–8.

[7]Chi Y, Wang YT, Wang Y, et al, (2012) Dry and noncontact EEG sensors for mobile brain computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(2):228–235.

[8] Cincotti F, Mattia D, Aloise F, et al, (2008) Non-invasive brain–computer interface system: towards its application as assistive technology. *Brain research bulletin* 75(6):796–803.

[9] Costanza E, Inverso SA, Allen R (2005) Toward subtle intimate interfaces for mobile devices using an EMG controller. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, pp 481–489.

[10] De Vos M, Gandras K, Debener S (2014) Towards a truly mobile auditory brain–computer interface: Exploring the p300 to take away. *International Journal of Psychophysiology* 91(1):46–53.

[11] De Vos M, Kroesen M, Emkes R, et al, (2014) P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier. *Journal of neural engineering* 11(3):2014.

[12] Debener S, Minow F, Emkes R, Gandras K, Vos M (2012) How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology* 49(11):1617–1621.

[13] Edlinger G, Krausz G, Laundl F, et al, (2005) Architectures of laboratory-pc and mobile pocket pc brain-computer interfaces. *Proceedings of2nd International IEEE EMBS Conference on Neural Engineering, 2005*.,pp 120–123.

[14] Gargiulo G, Bifulco P, Calvo R, Cesarelli M, Jin C, van Schaik A (2008) A mobile EEG system with dry electrodes. *Biomedical Circuits and Systems Conference (BioCAS) 2008.* IEEE, pp 273–276.

[15] Guan C, Thulasidas M, Wu J (2004) High performance p300 speller for brain-computer interface. *IEEE International Workshop on Biomedical Circuits and Systems, 2004, IEEE*, pp S3–5.

[16] Guger C., et al, P300 spelling device with g. mobilab and Simulink.

[17] Gupta S, et al, (2012) Detecting eye movements in EEG for controlling devices. *IEEE International Conference on Computational Intelligence and Cybernetics (CyberneticsCom),* 2012, pp 69–73.

[18] Gu ̈rko ̈k H, Nijholt A (2012) Brain–computer interfaces for multimodal interaction: a survey and principles. *International Journal of Human-Computer Interaction* 28(5):292–307.

[19] Hintermu ̈ller C, Kapeller C, Edlinger G, Guger C (2013) BCI integration: Application interfaces.

[20] H ̈ohne J, Tangermann M (2014) Towards user-friendly spelling with an auditory brain-computer interface: The charstreamer paradigm. *PLOS ONE* 9(6):e98,322.

[21] Hohne J, Schreuder M, Blankertz B, Tangermann M (2010) Two-dimensional auditory P300 speller with predictive text system. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2010, pp 4185–4188.

[22] Homan RW, Herman J, Purdy P (1987) Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology* 66(4):376–382.

[23] Huggins JE, Guger C, Allison B, et al (2014) Brain- Computer Interfaces, *In the Workshops of the fifth international brain-computer interface meeting: Defining the future.* 1(1):27–49.

[24] Jayadeva. Learning a hyperplane classifier by minimizing an exact bound on the VC dimension." *Neurocomputing* 149 (2015): 683-689, DOI http://dx.doi.org/10.1016/j.neucom.2014.07.062

[25] Jennrich RI, Sampson P (1977) *Stepwise discriminant analysis. Statistical methods for digital computers* 3:77–95.

[26] Keerthi SS, Chapelle O, DeCoste D (2006) Building support vector machines with reduced classifier complexity. *The Journal of Machine Learning Research* 7:1493–1515.

[27] Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews* 29(2):169–195.

[28] Lee TW, Girolami M, Sejnowski TJ (1999) Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation* 11(2):417–441.

[29] Li Y, Zhang J, Su Y, Chen W, Qi Y, Zhang J, et al, (2009) P300 based BCI messenger. *In proceedings of ICME International Conference on Complex Medical Engineering, 2009*.pp 1–5.

[30] Lievesley R, Wozencroft M, Ewins D (2011) The Emotiv EPOC neuroheadset: an inexpensive method of controlling assistive technologies using facial expressions and thoughts. *Journal of Assistive Technologies* 5(2):67–82.

[31] Lin CT, Ko LW, et al, (2008) Noninvasive neural prostheses using mobile and wireless EEG. *Proceedings of the IEEE* 96(7):1167–1183.

[32] Lin CT, Ko LW, Chang MH, et al, (2009) Review of wireless and wearable electroencephalogram systems and brain-computer interfaces–a mini-review. *Gerontology* 56(1):112–119.

[33] Lin MK, Nicolini O, Waxenegger H, Galloway GJ, Ullmann JF, Janke AL (2013) Interpretation of medical imaging data with a mobile application: a mobile digital imaging processing environment. *Frontiers in neurology*.

[34] Lin YP, Wang Y, Jung TP (2013) A mobile SSVEP based brain-computer interface for freely moving humans: The robustness of canonical correlation analysis to motion artifacts. *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2013, IEEE, pp 1350–1353.

[35] Lin YP, Wang Y, Jung TP (2014) Assessing the feasibility of online SSVEP decoding in human walking using a consumer EEG headset. *Journal of neuro-engineering and rehabilitation* 11(1):119.

[36] Mill ́an JdR (2013) Brain-computer interfaces. *Introduction to Neural Engineering for Motor Rehabilitation* 40:237.

[37] Mill ́an JdR, Renkens F, Mourino J, et al, Non-invasive brain-actuated control of a mobile robot. *Proceedings of the 18th international joint conference on Artificial intelligence,* (2003) pp 1121–1126.

[38] Nijholt A (2009) BCI for games: A state of the art survey. *Entertainment Computing-ICEC 2008,* Springer, pp 225–228.

[39] Nijholt A, van Erp JB, Heylen D (2008) *Braingain: BCI for HCI and games.*

[40] Nijholt A, Bos DPO, Reuderink B (2009) Turning shortcomings into challenges: Brain–computer interfaces for games. *Entertainment Computing* 1(2):85–94.

[41] Penghai L, DongDHWBM (2011) Research progress on application of brain-computer-interface in mobile peripheral control. *Journal of Biomedical Engineering* 3:038.

[42] Pfurtscheller G, Brunner C, Schl¨ogl A, Lopes da Silva F (2006) Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. *Neuroimage,* 31(1):153–159.

[43] Polich J (1997) On the relationship between EEG and P300: individual differences, aging, and ultradian rhythms. *International journal of psychophysiology* 26(1-3):299–317.

[44] Scherer R, Proll M, Allison B, Muller-Putz GR (2012) New input modalities for modern game design and virtual embodiment. *Virtual Reality Short Papers and Posters (VRW),* 2012 IEEE, IEEE, pp 163–164.

[45] Stopczynski A, Larsen JE, Stahlhut C, et al, (2011) A smartphone interface for a wireless EEG headset with real-time 3D reconstruction. *Affective Computing and Intelligent Interaction*, Springer, pp 317–318.

[46] Stopczynski A, et al, (2014) The smartphone brain scanner: A portable real-time neuroimaging system. *PloS one* 9(2):e86,733.

[47] Stopczynski A, et al, (2014) Smartphones as pocketable labs: Visions for mobile brain imaging and neurofeedback. *International Journal of Psychophysiology* 91(1):54–66.

[48] Stytsenko K, Jablonskis E, Prahm C (2011) Evaluation of consumer EEG device Emotiv EPOC. *Proceedings of MEi: CogSci Conference 2011*, Ljubljana.

[49] Takahashi K, Ohta T, Hashimoto M (2008) Remarks on EOG and EMG gesture recognition in hands-free manipulation system. In: Systems, Man and Cybernetics, 2008. SMC 2008., pp 798–803

[50] Tallgren P, Vanhatalo S, Kaila K, Voipio J (2005) Evaluation of commercially available electrodes and gels for recording of slow EEG potentials. *Clinical Neurophysiology* 116(4):799–806.

[51] Tonin L, Leeb R, Tavella M, Perdikis S, del Millan J (2010) The role of shared-control in BCI-based telepresence. *In proceedings of IEEE International Conference on Systems Man and Cybernetics* (SMC), 2010, pp 1462– 1466.

[52]Vernon S, Joshi SS (2011) Brain–muscle–computer interface: mobile-phone prototype development and testing. *IEEE Transactions on Information Technology in Biomedicine*, 15(4):531–538.

[53] Van Vliet M, et al, (2012) Designing a brain-computer interface controlled video-game using consumer grade EEG hardware. *Proceedings of Biosignals and Biorobotics Conference (BRC),* 2012 ISSNIP, IEEE, pp 1–6.

[54] Volosyak I, et al, (2011) BCI demographics II: how many (and what kinds of) people can use a high- frequency SSVEP BCI? *IEEE Transactions on Neural Systems and Re- habilitation Engineering*, 19(3):232–239.

[55] Wang YT, Wang Y, Jung TP (2011) A cell-phone-based brain–computer interface for communication in daily life. Journal of neural engineering 8(2):025,018.

[56] Wang YT, Wang Y, Cheng CK, Jung TP (2013) Developing stimulus presentation on mobile devices for a truly portable SSVEP-based BCI. *Proceeding of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 2013, IEEE, pp 5271–5274.

[57] Ward DJ, Blackwell AF, MacKay DJ (2000) Dashera data entry interface using continuous gestures and language models. *Proceedings of the 13th annual ACM symposium on User interface software and technology, ACM*, pp 129–137.

[58] Wei Q, Zou X, Lu Z, Wang Z (2014) Design and implementation of a mental telephone system based on steady-state visual evoked potential. *Journal of Computational Information Systems* 10(2):547–554.

[59] Williamson J, Murray-Smith R (2005) Hex: Dynamics and probabilistic text entry. *Switching and Learning in Feedback Systems, Springer,*pp 333–342 .

[60] Zhu D, Bieger J, Molina GG, Aarts RM (2010) A survey of stimulation methods used in SSVEP-based BCIs. *Computational intelligence and neuroscience* 2010.

MULTIMODAL ROBOTS AS EDUCATIONAL TOOLS IN PRIMARY AND LOWER SECONDARY EDUCATION

Gunver Majgaard
University of Southern Denmark, The Maersk Mc-Kinney Moller Institute, Campusvej 55, 5230 Odense M, Denmark

ABSTRACT

Multimodal humanoid robots have been used as educational tools in primary and lower secondary schools. The pupils involved were between 11 and 16 years old. The learning goals included: programming, language learning, ethics, technology and mathematics, e.g. practised by 7th grade pupils who programmed the robots and made the robots recite poems about the future. We conducted workshops for the teachers in didactical planning and programming of the robots. In the most successful settings, the pupils worked with academic objectives beyond programming and robotics. Through examples, we highlighted the potentials and the shortcomings in multimodal-robot-supported learning.

KEYWORDS

HCI, Human Robot Interaction, Design Education, Robots, Multimodal interfaces.

1. INTRODUCTION

In this article, we describe how humanoid NAO robots can be used as an educational tool in primary and lower secondary schools in Denmark. A NAO robot is a multimodal interface which uses touch, speech, gestures and eye gaze for interaction (Aldebaran Robotics, 2015). It is assumed that multimodal interfaces support more flexible, efficient and expressive means of interaction that are more akin to humans' experiences in the physical world (Sharp, 2007). And this is supposed to provide a richer and more complex user experience (Sharp, 2007). The paper contributes with an indicative example of how to use this technology in teaching and a summary of its educational multimodal properties. Multimodal interfaces have been used in primary and secondary education for many years in the form of LEGO Mindstorm, where children build and program mobile robots. The LEGO Mindstorm concept was inspired by Papert's (1993) ideas on creative and innovative learning (Resnick, 2009). The LEGO robots often look like futuristic vehicles. In contrast, the NAO robot has already been built and looks like a human being with arms, legs, body, and head. This provides a totally different approach. When you see the robot for the first time, you expect it to have some kind of humanlike behaviour (Kahn, 2007). As an educational tool, it provides the children with the possibility of exploring the design of multimodal human-robot communication. We give the schoolchildren in the project the possibility to design physical humanlike gestures and speech. We also prepare the pupils for a future, where robots might have prominent roles as social and assistive tools e.g. for people with disabilities. This provides a new and different perspective that has not been studied before in schools with normally functioning children. Maybe a multimodal humanoid robot motivates for learning and collaboration in a different way. Perhaps it gives rise to ethical discussions about robots' roles in society, in the future.

While working with the robots, the pupils receive initial insight and skills in the relationship between digital design, translation, symbolic coding and diagramming on the one hand and physical expression and communication on the other hand. According to Resnick (2009) digital fluency and literacy is an important learning goal in schools. The children should be able to produce interactive behaviour and not only react and consume others' interactive designs. This will provide a deeper understanding of the digital world. Blikstein (2013) even believe that digital fluency can have a democratizing effect because children are going to explore a technology that was previously controlled by experts only.

ISBN: 978-989-8533-38-8 © 2015

Children between 11 and 16 years of age used the robots in the classroom. There were about 24 children in a normal class setting and they shared 3 NAO robots. The teachers were initially on a two-day intensive introductory course in the technology and ICT-based educational design. The teachers then conducted experimental teaching for about eight to twenty hours.

The research question is: How can the multimodal NAO robots enrich children's learning? The methodical approach is qualitative, and in order to answer the research questions we collected lesson plans, evaluations, observations and in-situ interviews from the workshops and the classrooms. These empirical data are the basis for the examples and discussions in this article. The research methodology is based on design-based research, which is a research method suitable for studies of how technology and instructional design can support learning in the classroom (Majgaard, 2011).

The article is organised as follows: First, an introduction of robot technology in an educational context. Second, a theoretical section on how constructionism constitutes a theoretical basis for using robots in the classroom. Subsequently, we introduce the setting for the experiments and describe illustrative examples. These are related to the theory. At the end of each example there is a selection of the teachers' evaluations.

2. THE MULTIMODAL NAO ROBOT AS AN EDUCATIONAL TOOL

The NAO robot is a 58 cm tall humanoid developed by Aldebaran Robotics (2015), see figure 1(a) below. The NAO robot perceives the world through sensors, such as microphone, camera and tactile pressure sensors. And it communicates with the outside world by means of effectors, such as the motion of arms and legs through electric motors, sound and LED lights. The robot is programmed by a graphical block programming language, Choregraphe which is relatively easy to master for the novice, see figure 1(b).

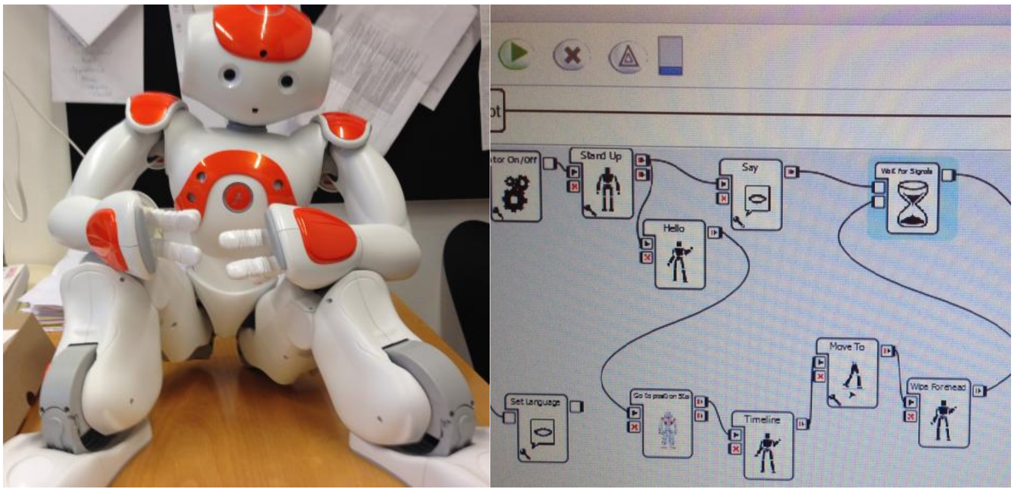


Figure 1. (a) A NAO robot; (b) Choregraphe programming environment

The robot is designed for use in education and research contexts and is currently used mainly in technical higher education and research environments. Pupils and research groups have, for example, developed interactive soccer-player behaviour into the NAO robots and enrolled them in a special RoboCup (2014). It is popular in the research field of human-robot interaction (HRI). The main goal of HRI is to enable robots to successfully interact with humans as they increasingly make their way into functional roles in everyday human environments such as homes, schools, and hospitals.

Other robotic concepts such as LEGO Mindstorm (2014) have been used in primary and lower secondary education where the pupils construct and program robots. Educational goals are related to innovation, experimentation, construction, electronics and programming.

Others have been using robots for language learning. Tanaka et al (2011) has been exploring different types of robots for foreign language learning. Latest she explored the use of a child-operated telepresence robot for the purpose of remote education. The robot was a medium for video conferencing between the pupils and a native English-speaking teacher in a remote destination (Tanaka et al, 2011).

Han (2005; 2009) has also been exploring home robots and robots as a teaching assistant in the field of language learning. In the case of the home robot Han explored the children's learning interests, concentration and academic achievements (Han, 2009). The robot delivered the content which was English dialogue for 6th graders. The results showed that the children were concentrated for a longer period of time and that the academic achievements and interest were higher using the home robot compared to web-based instruction and books with an audio tape. In the case of using robots as a teaching assistant in the classroom while learning English Han found that the children liked to relate to the robot.

Educational humanoid robots have been used as therapeutic tools for children with autism (Dautenhahn, 2007; Kozima, 2007). A popular example is tele-operated Keepon (Kozima, 2007) which was also used therapeutically for children with autism. The children were taught basic social skills such as eye contact and so-called joint attention. Social robots can motivate by creating new relationships and offer the children new social roles (Bertel, 2013). Kanda et al (2014) explored how robots can form long-term relationships with pupils. They developed the robot's behaviour so that it could recognise children, and the robot confided its personal matters to children who interacted a lot with the robot.

Humanoids have also been used as instructors for teaching, for example as a fitness instructor in a school setting, but the robot has a lot of motoric and interactive shortcomings (Nonaka, 2014). Our view is the opposite. We want the robot to become a partner – i.e. more than a mere object.. The children instruct the robot and evaluate the consequences.

3. LEARNING APPROACH - CONSTRUCTIONISM

How can pupils' learning abilities be stimulated by multimodal physically interactive educational tools such as NAO robots? To investigate this further, we looked back at the history of Papert's (1993) concept of constructionism. Papert was one of the first to combine physical interactive educational tools and learning theory. His thoughts built on Piaget's concepts of constructing cognitive schemes, based on the individual's interaction with the environment. According to Piaget, the learner constantly adapts his knowledge to new experiences. Papert believes that learning and physical interaction are linked, e.g. a child learns about construction while building a tower.

In Papert's perspective, learning takes place when developing physical or virtual productions, for example construction of a robot's behaviour. Papert further highlights the easy accessible programming languages as so-called "object to think with", where you get immediate feedback. Papert also emphasised that learning took place by solving problems and by developing an experimental approach to design processes (Papert, 1993).

4. RESEARCH METHOD AND THE DIDACTIC DESIGN

In this project, we use the previously mentioned design-based research, which is suitable for development of didactic design supported by technology (Majgaard, 2011; Van den Akker, 2006).

Structurally, the research process was divided into three phases: (1) Two-day workshop for teachers. The theme was hands-on technology activities during which two children from each class participated. Additionally, they developed didactical plans. (2) Teaching in the class ranging from for eight to twenty lessons and they had access to three NAO robots in that period. In some of the lessons the researchers participated as observers. (3) Teachers completed a questionnaire to evaluate the teaching.

Three schools attended each round of the workshop – so approximately nine school classes used the NAO robots. Lesson plans and evaluations can be seen in Danish on the project's Wikipedia page. The teachers also had access to each other's lesson plans and evaluations.

The questions asked in the written evaluations were part of the following categories: Educational goals; examples of activities; potentials; drawbacks; recommendations to other teachers, and achieved learning.

ISBN: 978-989-8533-38-8 © 2015

5. FINDINGS

The following section describes experiences from the workshops and an illustrative example from the teaching. Text bits in italics are quotes from the teachers' didactical plans.

5.1 Findings from the Workshops

On the first day of the workshop one or two teachers from each school participated in the event, each accompanied by approximately two pupils. This resulted in a few technical super users from each school. In addition, teachers could see how the pupils understood the technology, which they implicitly could use in their educational planning. The second day of the workshop had a didactic approach, and the teachers were presented with a didactic planning model. The model contained items on learning goals, activities, outcome, and organisation.

In their didactical plans, they defined goals such as*: "Foreign language - English: students talking in/using complete sentences. Focus on spoken English. Body language used as support for meaning and if you can't remember the word in English."* As an activity, they planned on working with "tongue twisters" e.g. she sells sea shells. The robot should recite the tongue twister and use supportive body language. In mathematics, they defined goals as: *"Mathematics: focus on oral mathematics and programming".* And they defined ethical learning objectives as well: *"Consider various ethical issues related to the use of robots in everyday life."* Learning in basic electronics: *"Fundamental understanding of circuits, components and programming."*

At the workshops, there was a tendency for teachers to initially formulate activities, and then articulate the learning goals. It might be a way for teachers to reflect on what the objectives of the activity are, and whether they are aligned with the overall curriculum.

5.2 An Overview and an Illustrative Example from the Classroom

A lot of different academic subjects and concepts were explored by the schools. Most of the teaching was multidisciplinary and combined disciplines such as programming and robotics in combination with English or Danish language teaching. The figure below summarises the subjects explored in the project.

.

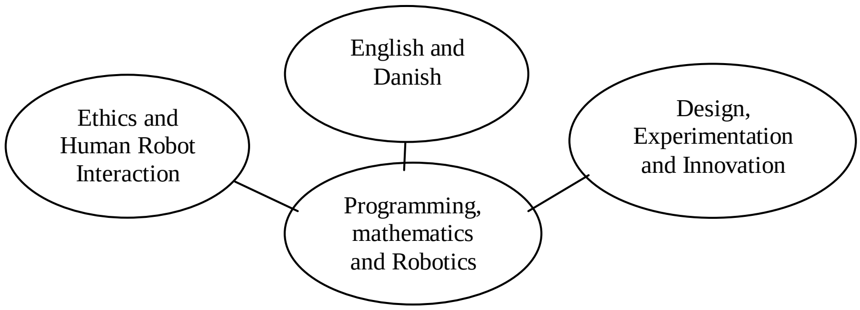


Figure 2. Overview of the academic subjects and concepts explored by schools

In programming, mathematics, and robotics they developed programming skills and got an initial understanding of sensors and effectors. In Danish, the pupils for example developed poems and robotic presentations. In foreign language learning of English, the pupils for example developed dialogues between the robot and users or robotic presentation of tongue twisters. The pupils also discussed ethical dilemmas such as the robot's role in everyday life in the future. Additionally, some of the pupils conducted real-life experiments in which they tested the robot in an everyday context such as how customers in the local grocery store experienced a talking robot, the robot as a fitness trainer or a dancer.

In the following, we introduce an example from a 7th grade school class, where 24 pupils worked for five weeks, two-four hours per week, with the NAO robot. The robot classes were run by two teachers. One teacher taught the pupils science and the other Danish. The first part of the process was carried out solely by the science teacher and provided basic knowledge and skills on how to program the NAO robot. The second part of the course was based on the first part. It was multidisciplinary and combined technology and Danish. The course ended with a presentation, where the robots recited and analysed poems written by pupils under the theme "future".

In the introductory part, a number of technical tasks in programming Choregraphe were carried out. They would get the robot to stand, dance, say self-chosen words in simulation mode, etc. Then they moved their applications to the physical robot and carried out the same tasks again, now in the physical world. Then they carried out activities, where the robot went into the adjoining rooms, avoided obstacles and turned its engines off, when it had carried out its activities. They worked with the robot's opportunities for physical animation using tactile programming, speech, and image recognition.

In the second part of the course, the activities circled around creating, analysing and presenting poems. The pupils worked in groups of four and each group implemented three types of presentations using the robot: (1) presentations of homemade poems which referred to a specific photo or picture, (2) self-selected poems which referred to specific pictures or photos, (3) analyses of their selected poems, and (4) analyses of their homemade poems. Technically, the robot walked towards a picture and pointed to it whenever it fitted into the presentation. During the entire course, two technically-minded pupils (who had also participated in the workshop) had a special responsibility for the robots. Other pupils had a responsibility for the computers, cables and so on.

In the following we discuss the illustrative example:

First: Cyclic repetition and learning depth. In a subsequent interview the Danish teacher emphasised that the pupils dived into the poems a second time after writing them, and they got the robot to present the poems in accordance with their ideas. They heard their own and others' poems several times. As they encoded the poem into the robot, they adjusted and expanded the poem. The teacher describes it as follows: "*they got more deeply into the subject matter".* In Papert's terminology, the robot was "an object to think with", when programming behaviour into the robot, the pupils saw how the robot responded. They then adjusted and refined the robot's behaviour.

Second: Orchestration of robot motion and time. Along with the encoding of the poem into the robot, they coupled physical movements. They took an active stance on how the robot should recite the poem, and the poem's content. Some had the robot sit while presenting and others experimented more actively with movements and gestures to support the recitation of the poem.

Third: Academic requirements led to synergy between technological and Danish academic immersion. After learning the most basic commands, they got an assignment which triggered their creativity and enthusiasm. There was established an academic and creative playing field in terms of the requirements for the final presentation. The clear requirements and objectives of the assignment gave the children a playing field where they could unfold. Through observation, we learned that the children used many facets of Choregraphe e.g. the digital animation.

In some schools, we observed that the children after having learned the basic commands – got an open assignment e.g. *"make an interesting experiment*" which they either completed quickly, got stuck in, or gave up on. Articulated goals and requirements, beyond getting to know the technology as in the example above, helped the children to unfold themselves academically and creatively.

6. HIGHLIGHTS FROM THE TEACHERS' EVALUATIONS OF THE NAO ROBOT

The following section presents quotes from teachers' evaluations of their teaching with NAO robots. Text bits in italics are quotes from the teachers' responses.

6.1 Motivation, Experimentation and "an Object to Think with"

The teachers were asked in a survey to articulate what made the NAO technology special in a school setting. They thought the robot itself was motivating in the beginning of the teaching process. *"The robots are in themselves very motivating for learning. They engage some of the children, who may not always be "very" concerned about school work. The trick is to find tasks that challenge the children to search out "academic" knowledge."*

The teachers highlighted the robot's opportunities to support children's active experiments, as it provided immediate feedback. *"It gives the opportunity to experiment. But some children experienced that they were more primitive than they had imagined ..."; "The robot's communication in spoken English was super."; "... The children were very motivated to use the robot's potential in terms of movement, speech, voice, recognition, etc.."* The robot responded immediately according to how it was programmed, which was not always the same as the children's ideas. This is comparable to Papert's description of "an object to think with", as this is one of the strengths of constructionist learning.

The teachers also expressed, what they thought worked well in their teaching. They highlighted that pupils quickly became self-propelled and that they had a good professional dialogue in the classroom, *"The pupils were quickly self-propelled*". There were good academic discussions amongst the pupils and between teachers and pupils. Other colleagues and pupils at the school were curious. A teacher expressed the following: *"It was a different way of teaching: learning rather than teaching."* This ability to be self-propelled may also be a result of pupils' interaction with the robot, in the development of the robot's behaviour, e.g. an aspect of constructionism's idea of experimentation, problem solving, and "an object to think with".

6.2 Danish, English, Ethics and Programming as Academic Themes

The teachers also drew attention to the positive link between Danish or English and the programming of NAO robot behaviour, including body language. One of the teachers wrote*: "The pupils' programming of the English tongue twisters worked really well. And they had to make the body language suited to what the NAO was talking about."*

In addition, the NAO robot puts focus on ethical dilemmas and real-life experiments with the robots. A teacher describes it like this: *"The ethical dilemmas worked really well, and the pupils felt that the discussions were interesting, and we saw a high degree of reflection regarding robots' influence on our future society. A group brought NAO to the local grocery store to see how other people would react to the presence of a robot, and if it was possible to get a dialogue going between the robot and the customers in the store. This group kept the motivation to work through the entire period and wanted to continue working with robots that interact with other people."*

Moreover, teachers described what they thought the children had learned. They featured programming and robotics skills: *"They have obviously learned to program." "The pupils have gained a greater understanding of robots functioning and applications. They related this to future dilemmas we will face as the technology gets better."*

Moreover, they highlighted the Danish and English academic skills with an emerging understanding of the supporting body language while presenting*: "I think the Danish technical terms and concepts rooted themselves better with the pupils. The pupils were very aware of the supporting use of body language." "English: Exercises about responding in complete sentences worked fine, but not as convincing as the supporting body language".*

A drawback in Danish was the robots' pronunciation. The pupils worked around this by spelling the words, so they fitted the Danish pronunciation. But then the spelling was not correct according to the Danish dictionary.

6.3 The Teachers' Recommendations: Clear Learning Objectives which go Beyond Getting to Know the Robot

Below is a selection of the teachers' recommendations. They emphasised in particular clear teaching objectives in addition to getting to know the robots. Moreover, they mentioned some technical problems, and the fact that the teacher must be familiar with the programming of the robot. A teacher explained it like this*: "Make sure to make the topic about more than robots."; "It is important that the teacher is familiar with the programming of robots. I think the hardest part was getting the robots to connect to the network. There were some network problems."*

To begin with, the NAO robot functioned more easily among older pupils. A teacher described it like this: *"The time aspect plays an obvious role. There was a long start-up, especially in the 4th and 5th grades. But the motivation makes the hours after the start-up super effective - compared to the outcome. (I conducted a* *small course with the 7th grade in IT electives where start-up clearly went much faster, and we quickly came to the important stuff)."* Another teacher wrote: *"We have worked with the NAO in the 9th-10th grades, and our use of community-related topics and ethics made the project exciting and educational for all pupils."*

7. SUMMARY

In the table below we summarise the findings based on observations and the teachers' feedback. The findings are divided into: Ways to learn; Interaction; Practical issues; Educational settings and didactics.

Table 1. Summary of multimodal properties found in the NAO robot as an educational tool.

Ways to learn: ➢ *"An object to think with"*. The robot provides immediate feedback to user during the development process and then becomes "an object two think with" (Papert, 1993). ➢ *Active experimentation and problem-solving*. The robot is suitable for active experimentation and problem solving, as the robot provides immediate feedback (Papert, 1993). ➢ *Self-propelled*. The children quickly became self-propelled in programming of robots. ➢ *Cyclic repetition and learning depth*. The children processed the subject matter in several rounds, which gave rise to a greater depth of learning. For example, the children adjusted their poems while coding it into the robot. **Interaction:** ➢ *Body language and robotic gestures*. The robot is suitable to support interaction using body language. ➢ *Dissemination*. The robot is very suitable for oral presentation using supportive body language. The robot has an easy to use text to speech function. ➢ *Affordance:* Form and expectation. Form and expectation must be closely linked in order to maintain the motivation. Because of the robot's muscular form some of the pupils explored the robot's potentials in the fitness area. E.g. one might think that the strong looking robot would be able to carry heavy objects – which it can't.	**Educational setting and didactics:** ➢ *Clear goals beyond getting to know the technology*. Exploring the "new technology" is not a full academic goal in itself. Additional academic goals in programing, mathematics or language are necessary. ➢ *Motivation.* The robot is motivating as an educational tool especially in the beginning. ➢ *It's faster and easier to introduce the robot for older pupils.* Children in 7th-10th grade worked more focused with the robots. ➢ *Multidisciplinary learning processes*. The robots were largely used in multidisciplinary disciplines e.g. programming and foreign language learning. ➢ *Academic requirements led to synergy between technological and language learning.* The requirements in the field of language learning made the pupils develop more complex programmes. ➢ *Organisation and structure*. Robots in the classroom sometimes presented a risk of chaos and turmoil. The experienced teachers countered this by structuring and organising the activities. **Practical issues** ➢ The robot may have difficulties to connect to the network due to local firewall settings and so on. ➢ Pronunciation is sometimes more phonetically correct rather than grammatically correct at least in Danish. ➢ *Problem with speech recognition* if the background is noisy.

8. SUMMARY AND CONCLUSION

In this article, we introduced the multimodal NAO robots as learning resources in the classroom. We investigated how technology can support and enrich the learning environment.

The pupils experienced both academic and technological benefits from the teaching. It was largely the constructionist way of learning that was the robot's strength. It became "an object to think with" as it immediately gave feedback in the development of programmes. The robot was used for teaching Danish, English, ethics, programming, and technology. The pupils particularly used the robot's text-to-speech and gesture features.

The teachers must be prepared for minor technical problems, such as network issues. Furthermore, three robots to 24 pupils are an absolute minimum.

Moreover, the two-pieced didactical plans were most successful. Piece one: getting to know the technology. Piece two: subsequent academic topics e.g. language learning with the robot as a lever for learning, more advanced programming and robot behaviour.

Be prepared to spend a couple of days to familiarise yourself with the technology and planning the course. There are no ready-made courses. But on our Wikipedia page you can locate individual course plans and evaluations from the study (Fremtek Wikipedia page, 2014).

REFERENCES

Aldebaran Robotics 2015. NAO Robot. https://www.aldebaran.com/en/humanoid-robot/nao-robot. Retriwved 2015-03-15.

Bateson, G. 2000. *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Chicago Press. ISBN 0-226-03906-4.

Bertel, L.B. & Rasmussen, D.M. 2013. "On Being a Peer: What Persuasive Technology for Teaching Can Gain from Social Robotics in Education." International Journal of Conceptual Structures and Smart Applications (IJCSSA), 1(2), s. 58-68

Blikstein, P. 2013. "Digital Fabrication and 'Making' in Education: The Democratization of Invention." FabLabs: Of Machines, Makers and Inventors. Bielefeld: Transcript Publishers.

Dautenhahn, K. 2007. Socially intelligent robots: dimensions of human–robot interaction. Philos Trans R Soc Lond B Biol Sci. 2007 April 29; 362(1480): 679–704. Published online 2007 February 13. doi: 10.1098/rstb.2006.2004 PMCID: PMC2346526

Han, J et al 2005.The Educational Use of Home Robots for Children. 2005 IEEE International Workshop on Robots and Human Interactive Communication.

Han, J. et al 2009. r-Learning Services for Elementary School Students. with a Teaching Assistant Robot. HRI'09, March 11-13, 2009, La Jolla, California, USA.

Kozima, H. & Nakagawa, C. 2007. A Robot in a Playroom with Preschool Children: Longitudinal Field Practice. Robot and Human Interactive Communication, 2007, s. 1058-1059.

Kahn, P., H. et al 2007. What is a human? Toward psychological benchmarks in the field of human-robot interaction. Interaction Studies 8:3 (2007), 363-390. ISSN 1572-0373. John Benjamins Publishing Company.

Kanda et al 2014. Poster from HRI 14

LEGO Mindstorm. 2014. http://www.lego.com/en-gb/mindstorms/ . Retrieved December 8, 2014

Majgaard, G., Hansen, J. J., Bertel, L. B. & Anders, P. 2014. Fra digitalt design til fysisk udtryk – anvendelse at 3-d-printere og NAO-robotter i folkeskolen (From digital design to physical expression – using 3-d-printers and NAO-robots). Mona. 2014, 4, s. 7-26

Majgaard, G., Misfeldt, M. & Nielsen, J. 2011. How Design-Based Research, Action Research and Interaction Design Contributes to the Development of Designs for Learning. p. 8-21. Designs for Learning 4/11. http://www.designsforlearning.nu/11/no2/DFL_02_11_majgaard_etal.pdf

Nonaka, I. & Takeuchi, H. 1995. The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford: Oxford University Press.

Papert, S. 1993. Mindstorms Children, Computers, and Powerful Ideas. Basic Books.

Resnick, M. et al. 2009. Growing up Programming: Democratizing the Creation of Dynamic. CHI 2009, April 4 – April 9, 2009, Boston, MA, USA ACM 978-1-60558-247-4/08/04

RoboCup. 2014. http://www.robocup.org/about-robocup/objective/ . Retrieved December 8, 2014

Sharp, H. 2007. Interaction Design: Beyond Human - Computer Interaction 2nd edition. Wiley.

Tanaka, F. and Ghosh, M. 2011. The Implementation of Care-Receiving Robot at an English Learning School for Children. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction HRI'11, March 6–9, 2011, Lausanne, Switzerland.

Tanaka F. at al 2014. Telepresence Robot Helps Children in Communicating with Teachers who Speak a Different Language. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction HRI'14, March 3–6, 2014, Bielefeld, Germany.

Van den Akker, J. 2006. Educational Design Research. Routledge.

TRANSITION CARDS: DESIGNING A CARD SORTING METHOD WITH AND FOR TEENAGE PATIENTS

Margaret Machniak Sommervold and Maja van der Velden
Department of Informatics, University of Oslo, Gaustadalléen 23 B, 0373 Oslo

ABSTRACT

This paper presents the process of designing and implementing a card-based method used in a Participatory Design process. The method was designed to support young patients in organizing and explaining their experiences and expectations surrounding their transition from pediatrics to adult healthcare. The paper argues in favor of not only including users in the design of new technologies, but also in the design of methods.

KEYWORDS

Participatory Design; young patients; transition; card sorting.

1. INTRODUCTION

During the course of our lives, we all undergo transitions. Transitions are processes or periods “in which something or someone undergoes a change and passes from one stage, state, form, or activity to another” (Fernandes & Landzberg 2004). Transition is a life altering part of our lives (ibid). Young people with chronic health challenges have their transition from puberty to early adulthood compounded by the challenge of moving from pediatric to adult medicine (Resnick & Bowes 2007). Transition in healthcare is being defined as “the purposeful, planned movement of adolescents and young patients with chronic physical and medical conditions from child-centered to adult-centered healthcare systems” (Blum 2002). A successful transition process is critical for ensuring patient compliance with future medical treatment (Alpay 2009). A significant amount of literature stresses the importance of good transitions (Baldassano et al. 2002; Alpay 2009; Fernandes & Landzberg 2004; Sawyer & Ambresin 2014; Campbell et al. 1996; Callahan et al. 2001) and presents the various problems that can occur during transition (Patterson & Lanier 1999; Kleinert 2007; Applebaum et al. 2013; Reiss & Gibson 2002). Despite the growing body of research, adolescents with chronic health challenges seldom make a smooth transition to adult healthcare (Patterson & Lanier 1999) and often have substantial, unmet needs (Bowen et al. 2010). While introducing technology is not always the answer, research indicates that ICTs hold the potential to support patients in various challenges related to their condition.

There is a lack of understanding of the teenage agenda and the challenges they face in transition – both as patients and adolescents (Patterson & Lanier 1999; Jacobson et al. 2001). There is also a lack of research involving teenagers (Fitton et al. 2012). While there is little research with teens, there is even less with teenage patients who are for good reasons ‘protected’ by ethical guidelines and restrictions. But while ethical guidelines and restrictions have been implemented to protect vulnerable populations, they hinder the inclusion of adolescents, which in itself is unethical (Lang et al. 2012). This is especially challenging in the context of Participatory Design (PD). Because of ethical and practical restrictions, vulnerable stakeholders’ views and experiences are excluded from the design solution (Robertson & Wagner 2012, p.73).

This paper presents the design and implementation of card sorting, which is a widely used method in Human Computer Interaction (HCI), and discusses how it helped to address the challenges of doing Participatory Design (PD) in a hospital setting with young patients. The paper focuses on i) the process of designing the cards, including a discussion of the underlying methodological and conceptual frameworks, ii) implementation of the card sorting exercise in a study mapping the transition experiences of young patients, and iii) what considerations must be taken when moving methods from PD into a hospital setting.

2. DESIGN AND IMPLEMENTATION

Although there is a scarce amount of research done with teenage patients, there exists a large amount of websites and mobile applications developed to support self-management and social connection between patients. Few of these initiatives have been designed together with patients, or aimed specifically at young patients. As Kanstrup et al. (2014) put it, "[p]atients tend to be human factors rather than human actors in the design of digital technology for healthcare" (ibid., p. 12). The authors interpret the absence of patient participation as an underestimation of patients' ability to contribute in design and innovation, and provide a list of studies demonstrating that given a chance, patients of different ages and capabilities can contribute to design and innovation of healthcare technologies. Mazzone et al. (2008) argue that teenagers are not only capable of productively contributing to design, but that they may benefit from the design process itself. Based on these arguments and our previous experience (van der Velden & Machniak 2014), we decided to take a PD perspective on transition and possible ICT solutions, because this perspective "takes into consideration the needs, interests and abilities of the youngsters, but also includes a more profound interest in their hopes, fears, dreams, and opportunities to express themselves as someone of importance" (Iversen & Smith 2012, p.113).

When implementing Participatory Design activities with teenage patients, two issues need to be taken into consideration. Firstly, teenage patients may identify first of all as teens, not as patients (van der Velden & El Emam 2013). Their needs and interests as teenagers may thus be different from their needs and interests as patients. Secondly, these teens may have cognitive, social, and/or physical challenges related to their diagnosis or treatment, which may limit their participation in design activities. Hence, the method we present in this chapter had to be designed with these considerations in mind.

2.1 Design of the Transition Cards

Card-based tools, such as the Card Sorting method (Hudson 2005), are popular design methods and serve different purposes and stages in the design process. Wölfel and Merritt (2013) provide an analysis of existing card-based design methods based on five dimensions: intended purpose of use, duration of use, methodology, customization, and formal/material qualities. Within the PD methodology, they found patterns cards being customized for a specific point during the design process. This was also the case in our study. The method we present was used early in the PD process as means of investigating real life practices and uncovering problem areas (Bratteteig et al. 2012). In order to examine the practices and needs of teenage patients, we decided to conduct card sorting with young patients. For the purpose of our work, we needed to design a new set of context-specific cards, which we called Transition Cards. Our work with the Transition Cards builds on previous work of (Culén and van der Velden, 2015), in which the authors designed and used Travel Experience Cards (TEC) to capture user experiences in public transportation. Inspired by Service Design Cards (Clatworthy 2011), the TEC card set was designed as a visual tool to both visualize a so-called customer journey with its service moments, such as checking departure times or buying a ticket, and to capture the travel experiences between those service moments.

The underlying motivation for the design of the Transition Cards was not visual methodology and service design, as in the case of the TEC cards, but rather an effort towards designing an age-appropriate tool for Participatory Design with young patients. In addition, when looking at transition as a process affected by and influencing the whole life of a patient, rather than a process taking place within a healthcare institution, operating with concepts such as touch points, or services does not provide a holistic picture of patients' experiences and expectations surrounding transition. The goal of the Transition Cards method was not to generate quantitative data for detailed analysis, but rather to aid participants in talking about transition.

2.1.1 Preparing the Content of the Cards

We started the design of the cards with a study of the main points and aspects surrounding transition. We reviewed published literature on transition, national guidelines for transition, and various tools for transition offered by healthcare institutions, such as checklists and other paper-based tools. We conducted a thematic analysis of the literature, and 4 themes emerged: people, things, skills, and feelings. Within each theme we tried to identify as many keywords and concepts as possible to give the participants a rich selection to choose from. Based on our previous research with teenage patients, we wanted the cards to represent both things related to being a patient as well as regular teenager, in order to encapsulate the holistic view of transition as something affecting all aspects of life as well as allowing the participants to present their stories not only as patients, but also “regular” teenagers (van der Velden & El Emam 2013).

Our main interest was not only to find out what or who was important during transition, but also to place these elements on a timeline. Therefore we decided that the cards should be sorted into time periods during the interviews. Brucker (2010) distinguishes between “open” and “closed” sort, where in open sort, the users sort the cards into meaningful categories, which they name themselves, while in “closed” sort the users are asked to fit each card into a predetermined category. Each of these approaches comes with its drawbacks and advantages. For the purpose of our study, we decided on a “closed” sort where the participants would sort the cards into time-periods, which were used to mark different stages in transition according to the Norwegian guidelines for transition. The categories we used were: *Child and Youth Clinic* representing the time between the age of 13 and 17, the number *18*, which was the recommended age for transition, and *Adult Medicine* which enveloped the period after transition. At that time, we believed these categories to be meaningful – both for the participants and for our analysis.

2.1.2 Anatomy of the Transition Cards

Based on our experiences with the TEC cards, we decided to use visual materials, such as pictures, in our cards. The use of images as representations augments cognition (Culén and van der Velden 2015). Images hold the potential of freeing the working memory and allow for “offloading of cognitive processes onto perceptual process” (ibid.). However, while images would provide sufficient information in the mapping of experiences in public transportation, visual representations of things related to transition could confuse the participants. Therefore we used images not as the primary source of information on the cards, but rather as illustrations for the key words and concepts written on the cards.

The images used on the cards were partly taken by us and partly taken from Internet sites. The images were then placed into Polaroid frames together with the keywords (Figure 1). The cards contained three pieces of information: a keyword, image, and a colored dot representing one of the four themes (*people, things, skills, feelings*), allowing for easier analysis after the interviews. The cards were printed out and laminated to allow for easy disinfection (van der Velden & Culén 2013).

Figure 1. Transition Card with the keyword “Doctor” (left) and a participant’s organization of the cards in categories

2.1.3 Recruitment and Participants

In total 17 young patients participated in this study, 11 girls and 6 boys, their ages between 13 and 22. We first applied and tested the method in a workshop with the Youth Council of the Akershus University Hospital (AHUS) in Norway. Five girls and three boys participated in the workshop. We then implemented the method during individual interviews with 8 patients recruited by the head of Research and Development of the Pediatric Clinic at AHUS, while they were receiving treatment at the clinic. Their ages were 13-18. We also conducted an interview with a young patient, 22 years old, who was recruited through a patient organization.

The Head of Research and Development of AHUS made sure that the participants were well enough to participate and ongoing consent was maintained throughout the whole interview session. The participants received a leaflet with information and a consent form, which was signed by the participants, researcher, and the guardians of participants under the age of 16. The participants were informed that their diagnosis would not be discussed, that they remained anonymous, and that they could retract their participation at any point in time. The study was approved by the Norwegian Data Authority and the Privacy Regulator of AHUS.

2.1.4 Workshop with the Youth Council and Pilot Interview

During the workshop at AHUS, we tested the cards with two groups of 4 participants. Each participant in a group received a set of cards with one theme (*people, things, skills or feelings*). They were then asked to choose three cards and place them under each category (*Child and Youth Clinic, 18, and Adult Medicine*, see Figure 2). Once the participants were done with a theme, they handed the deck of cards to the next person and received a new deck representing a different theme and were asked to repeat the exercise. This proved to be difficult for several reasons. First, as we thought the Transition Card sorting to be a collective effort, we operated with only one set of cards, which presented the next participant with a smaller selection of cards. Second, some of the participants had trouble choosing which cards they thought were most important for them. According to our observations, it seemed as the participants had problems with viewing the card exercise as a group process, which would result in showing their collective perception of transition. They seem to want to tell their individual stories. In their feedback, they advised us to implement the Transition Card sorting on individual basis or with maximum one more participant. The argument was that stories around transition and life with a diagnosis were so private and personal, that they didn't feel comfortable with sharing them with a larger group of people. Another comment was that choosing a limited amount of cards made it difficult to tell the whole story as they thought that it was important to convey that patients are also regular teens. In the evaluation of the workshop, the teenagers responded that they liked the cards and thought that this was a cool way to talk about transitions. The cards covered most of the things they wanted to talk about, but they wished to have separate categories for being a patient and a regular teenager. Since they could only choose a limited number of cards, they did not think that the result represented them as a patient *and* as a teenager.

Figure 2. Workshop with the Youth Council

After we received the feedback from the Youth Council, we re-evaluated our method and moved it from a group-based approach to individual approach. A pilot card sorting exercise was conducted with a young patient (girl 22), who had already transitioned to adult healthcare. Because of the lessons learned from the Youth Council, we asked the participant to use as many cards as she felt were relevant for her. This approach proved itself to be successful and the participant commented that she enjoyed using the cards. However, the categories we provided did not fit the participant's story. The participant decided to use the space above the category-cards as a space that covered both pediatrics as well as adult medicine. She explained that many of the things represented by the cards could be important in all of the stages during transition. This alteration resulted in better and broader descriptions of the cards by the participant. In addition, the participant did not use the "18" category other than commenting that it marked being legally an adult.

2.1.5 Final Adjustments and Implementation

After the feedback from the pilot exercise, we re-evaluated the categories and decided to continue with the categories suggested by the last participant, namely *Child and youth clinic*, *Adult Medicine,* and a new category *Both* which contained the first two categories. Because the Transition Cards method was to be implemented at the pediatric clinic while the participants received treatment, we had to re-think the physical qualities of the cards. The treatment rooms were not equipped with the large tables we had used during the previous sessions. In our earlier research, we used the "Cool Wall", which was a magnetic white board that we used for research on cool social networking sites and designing a health-oriented social networking site for teenagers (Machniak 2014; Culén & Velden 2013). Our experiences showed that magnets representing different sites and functions allowed us to work with bed-ridden patients and the whiteboard offered a great deal of flexibility during the interviews. Drawing on this experience, we produced a set of smaller-sized Transition Cards and applied magnetic tape to their backs. We drew the categories on the whiteboard and the participants could choose as many cards as they wished.

The majority of participants had no trouble with understanding the task and even made their own categories, showing that they understood the purpose of the cards and the categories on the board (see Figure 3). Two of the participants, however, showed some signs of confusion around the task, which proved that the Transition Cards sorting wasn't intuitive and self-explanatory, but rather dependent on careful explanation, especially with regard to the purpose and objective of the study. Once the purpose of the study was clear, the participants showed no signs of confusion and completed the exercise.

After sorting the four sets of cards, we asked the participants how and where they thought that ICT and other interventions could support them in transition. Because the participants had just categorized and talked about the various challenges and expectations related to transition, they could more easily target specific challenges and suggest how they could be supported in tackling these.

Figure 3. The final set of transition cards and the magnetic whiteboard (left) and an example of a participant altering the categories on the whiteboard (right).

3. DISCUSSION

Designing the Transition Card method showed itself to be a larger project than we had anticipated. Even with our experiences in working with teenagers, we acknowledge how fortunate we were to have an expert panel consisting of the members of the Youth Council who tested and gave us feedback on our method. Much of the literature on designing with future users and on card sorting as a method, suggest group activities as they result in more accurate and average representation of user needs. What we learned from the Youth Council was that talking about transitions or personal stories about managing one's condition in everyday life requires a high degree of sensitivity and confidentiality.

When working with vulnerable users (Culén & Velden 2013; Vines et al. 2014), we first and foremost need to strive toward protecting and respecting their integrity and privacy. When using card-sorting methods, Brucker (2010) suggests to advice the participants against overthinking the exercise. However, while rapid associations or 'quick and dirty' approaches might work when conducting the Card Sorting method, we cannot ask young patients to map their lives, including their needs, hopes, fears, and apprehensions, etc., by "just putting the card somewhere". Because the Transition Cards are meant to help them with envisioning their futures side by side with technologies, we need to allow them the time they need to carefully and meaningfully sort the cards into right place.

Our previous research indicates that young patients, teenagers in particular, tend to separate their identity as patients from their identity as young persons (van der Velden & El Emam 2013; van der Velden & Machniak 2014). Therefore it was important to meet the participants in a setting where they felt safe to talk from a patient perspective. Meeting patients in hospitals does however have its drawbacks. As the patients were receiving treatment, we were often interrupted by nurses who needed to check on the patients or adjust the medical equipment. The cards then served as a great way to pick up the interview just where we left it. The participants seamlessly continued on with the next card.

The Transition Cards fulfilled the criteria we mentioned in the beginning of this chapter. Because they were laminated, they could be easily disinfected. The magnetic board allowed the participants to sit or rest comfortably while using it and they were able to see the results in front of them. Although we had set the categories on the whiteboard, the participants could easily change these as they saw fit. The method allowed them to use only one arm. The visual appearance of the cards made the Transition Cards more attractive to use and several participants commented on them and thought they were nice and reminded them of Instagram pictures, which they considered *cool*.

The card sorting was split into sessions through the use of themes. The success of splitting the exercise into shorter sessions showed that many short activities with immediate output were appropriate for this age group as argued by Mazzone et al. (2008). The authors also recommend the tasks to be simple with defined objectives. In our study we found that the participants found the tasks easy, as long as they understood the objectives of the tasks. In addition, we found that the participants were more motivated and understood the task better when we provided a thorough description of the study as a whole, which indicated that they wanted to be well informed, not treated as merely informants, and contribute to the study in a meaningful way. When understanding that their contribution could help others in the same situation as their own, the participants also tended to think of other patients when sorting the transition cards, thus strengthening our claim that PD is a very appropriate methodology for young participants as it allows them to assume the expert position and establish themselves as someone of importance (Iversen & Smith 2012). This makes the whole experience more engaging and fun and strengthens mutual learning (Simonsen & Robertson 2012).

4. CONCLUSION

Participatory Design is a methodology to design technological artifacts as well processes that enable participants to engage in design activities (Simonsen & Robertson 2012). Methods, tools, and techniques need to be appropriated, altered, and made relevant for whatever participatory action we get involved in (Brandt et al. 2012). In a way, "[d]esign methods are like toothbrushes. Everyone uses them, but no one likes to use somebody else's" (Harrison & Tatar 2011, p.11). PD does not call for the creation of new, or the appropriation of existing, for the sake of more methods, but rather all methods need to fit the situation at hand (Bratteteig et al. 2012). While card-based methods are widely used and adapted in PD, descriptions of their design process and development are scarce.

In this project, due to the specific age group, life situation, and setting; the method and tools needed to be tailored and re-appropriated. What needed to be taken into account was: i) the challenges of designing with teenagers, ii) physical constrains (such as patient connected to an intravenous system or IV using only one arm, interruptions from the medical staff, need for disinfection), and iii) creating a method that was engaging and easy to understand and allowed the participants to express themselves in a meaningful way – both to the researcher and to themselves.

PD has traditionally been interested in finding tools for envisioning future ways of living through artifacts. Brandt et al. (2012) present how *enactment* helps people to envision themselves with new designs. Many methods such as scenarios, mockups, prototyping, etc. are being designed to enable future users to envision futures with new, unknown technologies. What is less researched is how future users envision not only new solutions, but also *their own* futures. In case of transition, our research shows that young patients don't know much about what transition entails or what will happen. Some of the teens we interviewed had simply not thought about it because they had some years left at the pediatric clinic, while other consciously avoided thinking about transition and the responsibilities their adult life would bring. Therefore several participants expressed that they found the Transition Cards helpful in structuring and organizing their thoughts and expectations. As one of our participants expressed *"Actually, I think it was great to have a lot [of cards] to choose from. Then you start to think about what the differences are, what you're going to miss, and what's going to be new. I feel that maybe it's good to have some cards to look at and talk about instead of just getting questions. Yeah, I don't know. It's just the way it is, but when you have a card to talk about – it gets better. So I think it's been great"* (girl, 17).

Because the participants could see their expectations and concerns in front of them, they could more easily suggest areas where they felt that ICTs could support them in transition, hence showing the Transition Cards to be helpful in envisioning both their own futures as well as technological solutions.

Through the process of designing the Transition Cards, as part of an effort to design ICTs supporting patients in transition, we learned that while it is important that technological solutions include the voices of the future users, but so should the design of the tools and methods in the design process. If we are to call our participants co-designers, we need to open up the process of designing the methods for their input. Especially in the context of vulnerable users or vulnerable life situations, it is important to consult the participants whether they think that the methods are adequate to elicit their contribution in terms of knowledge, experience, needs, etc.

This article presented a discussion and analysis of including users in the design of a participatory method. While this approach is still in its early stages, we hope that it can help designers and healthcare practitioners in learning more about the experiences and needs of young patients and that it will inspire further work in this field.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support and cooperation of the Children and Youth Clinic of Akershus University Hospital during the planning and implementation of the study. We thank all the participating teenagers for sharing their knowledge and experiences. We especially would like to thank the Youth Council for their willingness to spend time with us and share their expertise.

REFERENCES

Alpay, H., 2009. Transition of the Adolescent Patient to the Adult Clinic. *Peritoneal Dialysis International*, 29(Supplement 2), pp.S180–S182.

Applebaum, M.A. et al., 2013. Perception of transition readiness and preferences for use of technology in transition programs: teens' ideas for the future. *International Journal of Adolescent Medicine and Health*, 25(2), pp.119–125.

Baldassano, R. et al., 2002. Transition of the Patient With Inflammatory Bowel Disease From Pediatric to Adult Care: Recommendations of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. *Journal of Pediatric Gastroenterology & Nutrition*, 34(3), pp.245–248.

Blum, R.W., 2002. Introduction. *Pediatrics*, 110(Supplement 3), pp.1301–1303.

Bowen, M.E. et al., 2010. Health Care Transition in Adolescents and Young Adults With Diabetes. *Clinical Diabetes*, 28(3), pp.99–106.

Brandt, E. et al., 2012. Tools and techniques: Ways to engage telling, making and enacting. In *Routledge Handbook of Participatory Design*. New York: Routledge, pp. 145–181.

Bratteteig, T. et al., 2012. Methods: Organising principles and general guidelines for Participatory Design projects. In *Routledge Handbook of Participatory Design*. New York: Routledge, pp. 117–144.

Brucker, J., 2010. Playing with a Bad Deck: The Caveats of Card Sorting as a Web Site Redesign Tool. *Journal of Hospital Librarianship*, 10(1), pp.41–53.

Callahan, S.T. et al., 2001. Transition from pediatric to adult-oriented health care: a challenge for patients with chronic disease. *Current opinion in pediatrics*, 13(4), p.310.

Campbell, F. et al., 1996. Interventions to improve transition of care for adolescents from paediatric services to adult services. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd.

Clatworthy, S., 2011. Service innovation through touch-points : development of an innovation toolkit for the first stages of new service development. *15-28*.

Culén, A.L. & Velden, M. van der, 2013. The Digital Life of Vulnerable Users: Designing with Children, Patients, and Elderly. In M. Aanestad & T. Bratteteig, eds. *Nordic Contributions in IS Research*. Lecture Notes in Business Information Processing. Springer Berlin Heidelberg, pp. 53–71.

Culén, A.L. & Velden, M van der, 2015. Making Context Specific Card Sets – A Visual Methodology: Capturing User Experiences with Urban Public Transportation. *International Journal On Advances in Intelligent Systems*, in press.

Fernandes, S.M. & Landzberg, M.J., 2004. Transitioning the young adult with congenital heart disease for life-long medical care. *Pediatric Clinics of North America*, 51(6), pp.1739–1748.

Fitton, D. et al., 2012. Constructing the Cool Wall: A Tool to Explore Teen Meanings of Cool. *PsychNology Journal*, 10(2), pp.141–162.

Harrison, S. & Tatar, D., 2011. On Methods. *interactions*, 18(2), pp.10–11.

Hudson, W., 2005. Playing Your Cards Right: Getting the Most from Card Sorting for Navigation Design. *interactions*, 12(5), pp.56–58.

Iversen, O.S. & Smith, R.C., 2012. Scandinavian Participatory Design: Dialogic Curation with Teenagers. In *Proceedings of the 11th International Conference on Interaction Design and Children*. IDC '12. New York, NY, USA: ACM, pp. 106–115.

Jacobson, L. et al., 2001. How do teenagers and primary healthcare providers view each other? An overview of key themes. *British Journal of General Practice*, 51(471), pp.811–816.

Kanstrup, A.M. et al., 2014. Patient innovation: an analysis of patients' designs of digital technology support for everyday living with diabetes. *Health Information Management Journal*, pp.1–9.

Kleinert, S., 2007. Adolescent health: an opportunity not to be missed. *The Lancet*, 369(9567), pp.1057–1058.

Lang, A.R. et al., 2012. Not a minor problem: involving adolescents in medical device design research. *Theoretical Issues in Ergonomics Science*, pp.1–12.

Machniak, M., 2014. In pursuit of cool and its implications for the design process. In *Celebration, Transformation, New Directions*. CATAC'14. Oslo, Norway, pp. 66-82.

Mazzone, E. et al., 2008. Design with and for disaffected teenagers. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. NordiCHI '08. New York, NY, USA: ACM, pp. 290–297.

Patterson, D.L. & Lanier, C., 1999. Adolescent health transitions: focus group study of teens and young adults with special health care needs. *Family & Community Health*, 22(2), pp.43–58.

Reiss, J. & Gibson, R., 2002. Health care transition: destinations unknown. *Pediatrics*, 110(Supplement 3), pp.1307–1314.

Resnick, M.D. & Bowes, G., 2007. Us and them: worldwide health issues for adolescents. *The Lancet*, 369(9567), pp.1058–1060.

Robertson, T. & Wagner, I., 2012. Ethics: Engagement, representation and politics-in-action. In *Routledge International Handbook of Participatory Design*. New York: Routledge, pp. 64–85.

Sawyer, S.M. & Ambresin, A.-E., 2014. Successful Transitions: Beyond Disease Control to Better Life Chances. *Journal of Adolescent Health*, 54(4), pp.365–366.

Simonsen, J. & Robertson, T., 2012. *Routledge international handbook of participatory design*, New York: Routledge.

Van der Velden, M. & Culén, A., 2013. Designing privacy with teenage patients: Methodological challenges. In *Methods of Working with Teenagers in Interaction Design*. Workshop at CHI 2013.

Van der Velden, M. & El Emam, K., 2013. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. *Journal of the American Medical Informatics Association: JAMIA*, 20(1), pp.16–24.

Van der Velden, M. & Machniak, M., 2014. Colourful Privacy: Designing Visible Privacy Settings with Teenage Hospital Patients. In *ACHI 2014, The Seventh International Conference on Advances in Computer-Human Interactions*. pp. 60–65.

Vines, J. et al., 2014. Special topic: Designing for and with vulnerable people. *Interactions*, 21(1), pp.44–46.

Wölfel, C. & Merritt, T., 2013. Method Card Design Dimensions: A Survey of Card-Based Design Tools. In P. Kotzé et al., eds. *Human-Computer Interaction – INTERACT 2013*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 479–486.

PARTICIPATIVE DEVELOPMENT OF TOUCHLESS USER INTERFACES: ELICITATION AND EVALUATION OF CONTACTLESS HAND GESTURES FOR ANESTHESIA

Daniel Künkel[1], Birgit Bomsdorf [1], Rainer Röhrig[2], Janko Ahlbrandt[3] and Markus Weigand[4]
[1]Hochschule Fulda - University of Applied Sciences
[2]Carl von Ossietzky Universität Oldenburg
[3]Universität Heidelberg
[4]Justus-Liebig-Universität Gießen

ABSTRACT

Methods for developing touchless user interfaces are still under investigation. Promising approaches to come up with usable gestures combine two phases: First, gestures are elicited from users by asking them to demonstrate gestures for given tasks or dialogs. Afterwards a gesture set is extracted and tested with users in a subsequent evaluation phase. The steps within the two phases may differ due to different challenges to overcome, like the gesture extraction. This paper reports on the methods and results of those two phases in the context of interactions with a medical information system during anesthesia induction.

KEYWORDS

Gesture Elicitation, Gesture Evaluation, Contactless Hand Gestures, 3D-Gesture Interaction, Anaesthesia Information Management Systems, Operating Room Information Systems

1. INTRODUCTION AND BACKGROUND

Contactless gesture interactions become more and more popular. Users don’t need to hold a device in their hands (e.g. the Wiimote Controller) and the industry offers mature products for everyday use (e.g. Leap Motion or Kinect). Researchers also benefit from this trend. It will open up new possibilities in the field of Human Computer Interaction. After touch-based gesture interactions have been established, the industry sees touchless gestures as the next logical step. In the field of medicine, gesture interactions appear as particularly interesting, because of the assisting intent, sterile hands or the doctor-patient relationship. Interactions by contactless gestures (3D-Gestures) are used not only in games but more and more in so-called serious applications, e.g. for rehabilitation (Camurri et al. 2003) or browsing of radiology images (Wachs et al. 2008). For the work presented below the context of anesthesia was chosen as application domain.

Gesture controlled user interfaces have been investigated for several years. Developing systematically intuitive and ergonomic 3D-gesture interactions, however, is still challenging. Different concrete procedures for designing gestures are reported in literature that can be classified according to different gesture development criteria:

- **Gestures’ origin** indicates whether they are user-elicited or designer-determined. In the case of user-elicited gestures users are asked to demonstrate gestures, e.g. in (Rapp and Strube 2002), that are collected, e.g. videotaped, and analyzed afterwards. Gestures originating from the design team, e.g. in (Gerling et al. 2012), are determined based on the designers’ knowledge and experience as well as on literature, guidelines or established gestures (e.g. pinch or zoom gestures). The two approaches, eliciting and determination, may also be combined, such as in (Rupprecht et al. 2013). Both user-elicited gestures and designer-determined gestures are usually evaluated through user tests before being implemented.
- A further differentiation is related to **technical limitations** that impact the resulting gestures. The authors in (Nielsen et al. 2004) contrast technology-based with human-based approaches. In the first of the two, gestures are implemented before being evaluated. Identified gestures are constrained by current

technology resulting in solutions that may be undesired from the user perspective. Modifications are costly if at all practicable. In human-based approaches gestures are user-elicited and subsequently implemented. This time, however, users may want gestures that are not realizable by up-to-date technology but the chances are that the gestures are easy to perform and to remember, intuitive, metaphorically and iconically logical combined with the functionality, and ergonomic in terms of physical effort (Nielsen et al. 2004).

- A **mixed human- & technology-based** approach combines the two. In (Bomsdorf and Blum 2014) users demonstrate gestures in front of a Kinect sensor and based on a given gesture recognition. Hence, technical limitations are taken into account but gestures are still user-elicited – and currently realizable at the same time. A further mixed human- & technology-based approach is the theater-system technique (Mahr et al. 2011) that combines eliciting and determination of gestures. Gestures are predefined but selected parameters (e.g. the radius of circle gesture) are adjusted by users performing the gestures. Hereby predefined gestures are evaluated and appropriate parameters for subsequent gesture implementation are collected within the same step.
- The criteria **given input** distinguish whether only functions of the target system or prototypes showing UI presentations are taken into account (Top-Down vs. Bottom-Up development) (Nielsen et al. 2004).

All the mentioned approaches realize a user-centered design process, at least when it comes to the gesture evaluation. The criteria can be used to indicate the degree of user participation. In this paper a study to gesture development is presented. The applied methods followed the concept of rapid and low-fidelity prototyping with high user involvement. The focus of the study was to put on three selected user actions and thus to three gestures. Hereby the development effort was reduced in this first study since not only gestures but also the methodical approach are under development. The next section introduces important aspects of an anesthesia workplace. Then the methodical approach and the main results of the study are presented. The paper concludes with a discussion and future work.

2. STUDY

2.1 Context of Use

This study was conducted at a department for anesthesiology and intensive care medicine of a university hospital (UKGM-GI). The goal was to develop gestures to interact with the Anesthesia Information Management System (AIMS) NarkoData® (IMESO GmbH, Germany) (Branitzki et al. 2007). The AIMS collects and displays anesthesia-related data during the pre-, intra- and postoperative stages. While most data is automatically collected from attached medical devices, such as anesthesia machines and patient monitoring systems, some have to be entered manually using a mouse and a keyboard.

Among manually entered data points are specific *timestamps* within the perioperative process, as defined in (Bauer et al. 2008). Three specific points in time during an endotracheal anesthesia were chosen from the timestamps for the study presented in this paper: the beginning of a patient's anesthesia (start anesthesia induction), the end of intubation process (intubation) and anesthesia release time of the patient (anesthesia ready).

The setup of anesthesia workspaces varies a lot. Figure 1 depicts a representative setting of anesthesia workspace at the UKGM-GI. Adhering to evidence based guidelines, the anesthesia induction is performed by an anesthesia specialist physician (A), and a nurse anesthesist (B). The workplace comprises of different anesthesia devices, especially anesthesia machines and patient monitoring devices, usually positioned near the patient (C). The computer running AIMS client software is of lower priority and may therefore be out of a suitable range to be used. In the setting of Figure 1 the computer is on the far right (currently used by person A). The mandatory position of the anesthesia specialist is at the head of the patient. In some cases, the AIMS computer is even positioned behind the anesthesia specialist's back. Hence the timestamps are documented asynchronously and from memory. Thus, they are not perfectly accurate leading to imprecise documentation. The delay between action and documentation (of *time-critical timestamps*) was the main reason to investigate gesture interaction in this setting - besides hygiene demands. Interacting by means of gestures the documentation could be performed at the actual point in time – eliminating the delay.

Figure 1. Exemplary setting of an anesthesia workplace (panoramic shot)

2.2 Methodology

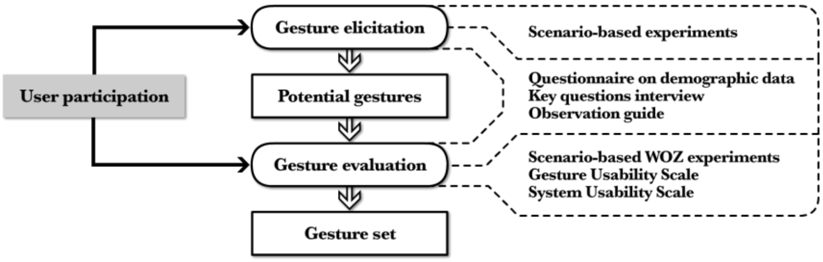


Figure 2. Overview of the applied gesture study.

The study consisted of two main phases (see Figure 2). The elicitation phase started with gathering user demonstrated gestures by means of scenario-based experiments such as in (Rapp and Strube 2002; Rupprecht et al. 2013; Nielsen et al. 2004). Subsequently from all user-elicited gestures such gestures were extracted that could be part of a solution. These potential gestures were investigated further within the evaluation phase. This time, scenario-based Wizard of Oz (WOZ) experiments were carried out, within which the well-known System Usability Scale (SUS)[1] and a Gesture Usability Scale (GUS) developed by the authors were used for measuring the usability. The questionnaires on demographic data, the questionnaires guiding the user interviews, and the observation guide were nearly identical to both the elicitation and the evaluation. In addition, the scenario descriptions and the video recording setting were very similar in the two phases.

Task analysis yielded that timestamps are not set in a uniform order. Sequences may vary, for example, between operating rooms and intensive care units. The goal, therefore, was to develop semaphores (an alphabet of gestures) (Quek et al. 2002) rather than a simple “next time-marker” gesture.

2.3 Settings

The parts of the study involving anesthesiologist were conducted on the job and at a real anesthesia workplace (comprising a computer with the AIMS software, an anesthesia artificial ventilator and a patient monitoring system). The interruption time caused by the study had to be reduced as much as possible. The human patient was substituted by a dummy (C) that usually is used in medical training. Each participating anesthesia specialist (A) stood at the head of the dummy and was assisted by a medical specialist (B) during a simulated endotracheal anesthesia. However, B did not interact with A taking part only to provide an authentic situation as much as possible. The gestures were only videotaped, i.e. no motion sensing input device was used not to be technically restricted. Two camcorders were in action during the elicitation, but only one within the evaluation since in this case different perspectives on gesture performance were less important than in the elicitation phase.

The evaluation was conducted as a Wizard of Oz (WOZ) experiment. WOZ are frequently applied to gesture eliciting as well as to their evaluation (Rupprecht et al. 2013). The advantage is that gestures do not have to be implemented beforehand (in contrast to technology-based approaches) since a hidden person, the

[1] System Usability Scale (SUS) – An Improved German Translation of the Questionnaire. [Online] Available from: http://minds.coremedia.com/2013/09/18/sus-scale-an-improved-german-translation-questionnaire [Accessed: 1st June 2015].

so-called wizard, emulates the functionality of the system providing users with feedback. WOZ was not used within the first phase of the study presented here since the gesture elicitation did not necessitate feedback. It was important during the evaluation to allow the user to distinguish right from wrong gesture execution.

Due to lack of space and the different workspace settings (see chapter 0), visual feedback within a gesture interaction was not suitable. Thus it was decided to indicate system reactions by sounds, since spoken language is reserved for important communication (among medical staff). Furthermore, to avoid excess noise, discrete gestures are used triggering a system reaction immediately after a gesture is completed, instead of providing feedback during its entire performance as in (Ruiz et al. 2011).

Based on these considerations, a specific sound was created for each gesture acknowledgement, and additional one to indicate errors. Each feedback was presented immediately after gesture execution and triggered by the wizard through the WOZSound application that was purpose-built for this test. The wizard was visible for the participant due to the lack of an appropriate, separate room, but the technique running WOZSound was hidden from them. Although the AIMS was not really used, even not simulated by the wizard, the participants believed they are interacting with that system while setting time markers by gestures.

2.4 Participants

The elicitation was carried out with 4 female and 8 male anesthesiologists aged from 27 to 47. They answered that they work between 5 and 48 hours per week in anesthesia, and have 0 to 20 years experience. The participant group consisted of 4 medical specialists and 8 resident physicians.

Ten participants took part in the evaluation (aged from 28 to 58, 4 female and 6 male. They worked between 16 and 42 hours per week, with 1.5 to 31 years work experience. This time the group consisted of 4 medical specialists, 6 resident physicians. All participants of the evaluation were different from those who were involved in the elicitation.

In both the elicitation and the evaluation all participants were experienced with touch interfaces (e.g. Smartphones, tablets, multitouch mouse and trackpads) and use them intensively. In contrast, only 10 of 22 participants (5 in elicitation, 5 in evaluation phase) have used contactless hand or body gestures (Kinect, Wiimote, Leap Motion Controller).

2.5 Gesture Elicitation

2.5.1 Scenario-based Experiments

The participants subsequently performed two similar scenario-based tasks for executing the simulated endotracheal anesthesia. Each task consisted of three actions, one per selected timestamp. In the first scenario, the participants recorded the times as usual with mouse and keyboard.

- Scenario I.1: Perform an endotracheal anesthesia with Mrs. Eichinger. You start with the first injection of the anesthetic. Perform all further necessary steps and document the time markers *start anesthesia induction, intubation* and *anesthesia ready* with the procedure you know.

Hereby the hypothesis of time-critical timestamps were checked and actually confirmed for this group. In the second scenario, the participants demonstrated gestures to document the times.

- Scenario I.2: Mr. Miller has also to be prepared for surgery by means of an endotracheal anesthesia starting in the same way as for Mrs. Eichinger and setting the three time markers *start anesthesia induction*, *intubation* and *anesthesia ready*, each with a gesture of your choice. Please tell us your association with each gesture.

Afterwards they were asked about their willingness to use gestures in anesthesia on a five-point Likert scale (1 to 5, 1 means *not at all*, 5 means *definitely*). The personal willingness of the participants were 4.17 on average; estimation of colleagues' willingness was scaled down 3.58 on average. All but one of them prefer to use gestures instead of mouse and keyboard.

2.5.2 Gesture Extraction and Documentation

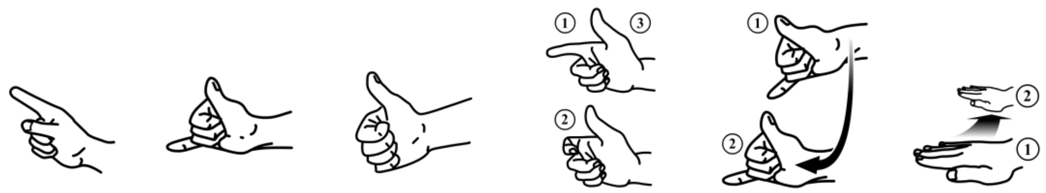
The user-suggested gestures were recorded from two perspectives resulting in two videos per gesture and participant. Particularly in more extensive studies (more gestures, more participants) the amount of data would not be easily manageable. Gestures notations such as ASCII Stokoe (Mandel 1983), HamNoSys (Hanke 2004) or Labanotation (Guest 2005) are not only used for documentation but also for analysis purposes. They are, however, difficult to learn because of their comprehensive set of symbols, meanings, and combination rules. A transcription would have been very time-consuming. Other approaches to gesture documenting utilize text and images (e.g. Nielsen et al. 2004; Rupprecht et al. 2013).

Figure 3. Exemplary dynamic gesture description from the perspectives front (upper) and side (lower)

One image per gesture as in (Nielsen et al. 2004; Rupprecht et al. 2013) is often not appropriate to represent movements and may affect the extraction of gesture vocabulary. Therefore, in this study dynamic gestures were described by image sequences (see Figure 3) to get rid of unnecessary video information. The image sequences were produced manually by selection of the most important key frames taking into account the association of the participants. The number of sequence images varied depending on the length and complexity of each gesture. The synchronous video recordings taken from the two perspectives enabled a parallel presentation of the image sequence of a gesture performance. Thereby the dimensions of gestures became clearly visible. Poses, however, were still shown by single frames. Blurred images were sorted out at least.

2.5.3 Potential Gestures

The 12 participants suggested 37 gestures (one person gave 2 gestures for intubation). Among them, they demonstrated 30 one-hand gestures, but only 7 two-hand gestures. Dynamic gestures were used 14 times, whereas in total 23 static gestures were shown. 24 gestures were significantly different and only a few were similar per process time. Therefore, gesture extraction resulted in 5 different gestures for the process time anesthesia beginning, 13 for intubation, and 8 for anesthesia released.



a) Index finger b) Hand tubus c) Thumbs up d) Shot e) Tubus roll f) Swipe

Figure 4. Static and dynamic gesture set illustrations

Two different potential gesture sets were created to check whether the participants prefer static or dynamic gestures. The associations of the participants were, besides gesture distinctness, a main criteria in the gesture extraction. The result is shown in Figure 4. The sketches (a) to (c) denote the static gestures for anesthesia begin (an index finger posture was associated with “first step is done and communicated to the colleague”), intubation (hand tubus posture associated with the device) and anesthesia release (thumbs up posture indicating ok, everything is prepared). The sketches (d) to (f) represent the dynamic gestures, whereby the numbers denote the order of single steps of gesture performance. The shot gesture (d) was associated with the injection at anesthesia begin. The intubation gesture (tubus roll gesture) inspired by the typical downward and roll motion while inserting a tube. The release gesture (swipe gesture) originated from the movement of taking the patient away into the next room.

ISBN: 978-989-8533-38-8 © 2015

2.6 Gesture Evaluation

2.6.1 Scenario-based WOZ-Experiments

Again, based on the scenario from the elicitation phase (see chapter 0.1), two analogue task scenarios were used. Both the static and the dynamic gesture sets were evaluated by each participant. The set to start with was randomized. The wizard demonstrated the gestures of the "first" set. Afterwards the participant performed the gestures a few times and the feedback sounds were introduced. Subsequently the other gesture set was trained as before to be used within the "second" scenario. In the two scenarios error cases were inserted to observe reactions to it. The error sound was played back after a defined gesture even though it was performed correctly. The failure was indicated in the first scenario after the release anesthesia gesture and in the second scenario after the intubation gesture.

After each scenario run the participants evaluated all single gestures by means of the GUS (see next paragraph). After the second scenario they additionally filled out the SUS, and were also asked about the appropriateness of the given interaction space and feedback sounds.

2.6.2 Rating of Gestures

The Gesture Usability Scale (GUS) (see Appendix) was developed for this study, by which the users assessed the gesture usability. The GUS is based on the structure of the SUS. It includes 10 items that were derived from the Human Interface Guidelines by Microsoft[2] and by Intel[3] and capture the aspects effectiveness, learnability, variability, distinctness, and fatigue of gestures. Each item is rated on a five point Likert scale from 0 = *strongly disagree* to 4 = *strongly agree*. A gesture's score value is calculated according to the SUS (between 0 and 100), i.e. the higher the score the better the gesture usability is.

The average score values of the single gestures were very similar (ranging from 79.6 to 87.5), i.e. the gesture usability was very high. Half of the participants preferred the static and the other half the dynamic gestures, although it was mentioned several times that postures were easier to perform, since they only have to be held for a short time. A dependency between the GUS ratings and a preferred gesture set was not observable based on a quantitative analysis. For example, the static gestures preferred by participant 7 were scored lower than the dynamic gestures. All in all, the evaluation showed no preference for one of the two gesture sets.

2.6.3 Further Results

The data collected during observation showed that in a few cases participants did not remember a gesture (thumbs up, hand tubus, and shot, each one time) and thus were re-demonstrated by the test manager. One person performed a wrong gesture (tubus roll). Three participants hesitated but executed the right gesture (again thumbs up, hand tubus, and shot, but different person). In faked error cases 3 of the 10 participants did not react to the error sound during the first scenario. They performed the same, right gesture a second time after the error was pointed out to them. In the subsequent scenario they immediately re-performed the gestures after the error sound. Participant 5 and 8 disliked to perform the shot gesture near a still awake patient. Participant 10 had problems with the finger coordination for the static and dynamic intubation gestures.

The SUS was taken to rate not only the gestures but the whole study prototype that was composed of the AIMS (from the participant point of view), anesthesia artificial ventilators and a patient monitoring system, the potential gestures, and the auditive feedback. The prototype was evaluated with 84.25 points, i.e. it was ranked to be acceptable (fourth quartile) with an excellent to good score. Asked about alternative forms of feedback, 7 participants favored light signals and other forms of audio. Furthermore, the intubation feedback was inappropriate since it sounds like slurping and not, as intended, like absorbing.

---

2 Microsoft® Developer Network. Kinect for Windows Human Interface Guidelines v1.8.0. [Online] Available from: http://msdn.microsoft.com/en-us/library/jj663791.aspx [Accessed: 1st July 2015].

3 Intel® Perceptual Computing SDK. Human Interface Guidelines [Online] Available from: https://software.intel.com/sites/default/files/article/401008/perc-humaninterfaceguidelines.pdf [Accessed: 1st July 2015].

3. DISCUSSION AND FUTURE WORK

The methodical approach has proven to be valuable. The target group was involved from the beginning following a user-centered approach: A contextual task analysis was performed prior to the study, by which useful information was gathered. The subsequent gesture development was dominated by user participation. Gestures were user-elicited implementing a human-based approach, that was only restricted by the pre-defined interaction space but not by technical constraints. Furthermore, it was a bottom-up approach. In the chosen application domain of anesthesia, to the authors' knowledge, no comparable work exists. Hence, there are no requirements or hints a technical prototype or implemented gestures (such as for a technology-based approach) could have been based on.

The elicitation process yielded an appropriate number of gestures for the specification of potential gestures. The workflow for extracting useful gesture representations from the captured sequences were developed by the authors using existing technology. The representations served the developers to get insights of the movements and poses within the 3D space. Their combination with the participants' association was valuable in defining the sets of potential gestures.

The GUS was applied in this study for the first time and will be improved in the context of the follow-up study. The GUS data, in contrast to the other data collected during the WOZ-experiment, allowed limited statements only. A concrete classification as in the case of the SUS, is currently not possible. A significant tendency for one of the evaluated gestures set was not verified. In future studies, the GUS but an updated version seems to be a promising approach to gesture evaluation with larger samples.

The study was a preliminary study and therefore reduced to 10 and 12 participants, respectively, and to only three timestamps. Thus, there's a lack of significance. Larger sample sizes, especially for the evaluation phase, would have been more valuable to find a gesture set for the given task and context. In the task analyses, however, a broader set of process times have been considered because in the intended follow-up study further gestures will be developed. In that study also mixed static and dynamic gesture sets will be investigated. In a proof-of-concept implementation, we've selected static gestures, which were easily implemented in a prototype called *GestureSimulator* using the leap motion controller. Because of technical constraints we had to adapt some of the gestures, e.g. the static hand tubus gesture (see Figure 4.b) has been rotated 90 degrees to the left (lying hand tubus).

The feedback sounds were not derived methodically but we are aware that the selection of the sounds may have an impact. This will be more investigated in a follow up study. All in all, there was big interest in contactless gestures interaction as well as in further investigation of the process in the domain of anesthesia.

REFERENCES

Bauer, M. et al, 2008. Glossar perioperativer Prozesszeiten und Kennzahlen – Eine gemeinsame Empfehlung von DGAI, BDA, BDC und VOPM. *Anästh Intensivmed*, *49*, pp. 93-S105.

Bomsdorf, B., & Blum, R., 2014. Early Prototyping of 3D-Gesture Interaction within the Presentation-Gesture-Dialog Design Space. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques* (pp. 12-23). Springer International Publishing.

Branitzki, P. et al, 2007. Spezielle Empfehlungen und Anforderungen zur Implementierung eines Anästhesie-Informations-Management-Systems. *Anästh Intensivmed*, *48*, pp. 282-290.

Camurri, A. et al, 2003. Application of multimedia techniques in the physical rehabilitation of Parkinson's patients. *The Journal of Visualization and Computer Animation*, *14*(5), 269-278.

Gerling, K. et al, 2012. Full-body motion-based game interaction for older adults. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1873-1882). ACM.

Guest, A. H., 2005. *Labanotation: the system of analyzing and recording movement*. Psychology Press.

Hanke, T., 2004. HamNoSys-representing sign language data in language resources and language processing contexts. In *LREC* (Vol. 4).

Mahr, A. et al, 2011. Determining human-centered parameters of ergonomic micro-gesture interaction for drivers using the theater approach. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 151-158). ACM.

Mandel, M., 1983. ASCII-Stokoe notation: A computer-writeable transliteration system for Stokoe notation of American Sign Language.

Nielsen, M. et al, 2004. A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *Gesture-Based Communication in Human-Computer Interaction* - 5th International Gesture Workshop (pp. 409-420). Springer Berlin Heidelberg.

Quek, F. et al, (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *9*(3), 171-193.

Rapp, S., & Strube, M., 2002. An Iterative Data Collection Approach for Multimodal Dialogue Systems. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)* (pp. 661-665).

Ruiz, J. et al, 2011. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 197-206). ACM.

Rupprecht, D. et al, 2013. Towards a Gesture Set for a Virtual Try-On. In: *Proceedings of IADIS International Conference Interfaces and Human Computer Interaction*, IADIS MCCSIS, IADIS Press

Wachs, J. P. et al, 2008. A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association*, *15*(3), 321-323.

APPENDIX: GESTURE USABILITY SCALE

	Strongly disagree				Strongly agree
1. I think that I cannot remember this gesture	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
2. I would imagine that most people would learn to use this gesture very quickly	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
3. I think that this gesture is not different enough from the other gestures	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
4. I think that the gesture differs from normal movements at work and therefore should not be carried out incase of mistake	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
5. I think the feedback does not match the gesture	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
6. I felt very secure in using this gesture and did not hesitate when executed	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
7. I think that the gesture does not fit the function	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
8. I think I could work faster with this gesture than with a mouse and keyboard	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
9. I am very overwhelmed by the gesture	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
10. I think that I would like to use this gesture frequently	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

A REVIEW OF HCI PATTERN TOOLS

Jürgen Engel, Christian Herdin and Christian Märtin
Augsburg University of Applied Sciences, Faculty of Computer Science
An der Hochschule 1, 86161 Augsburg, Germany

ABSTRACT

The construction of user interfaces for interactive systems typically requires the skills of software developers and HCI specialists who need to cooperate intensively with platform and marketing experts in order to arrive at solutions with high levels of software quality, usability, and user experience. Concerning this matter we develop a framework which combines model-based user interface development practices with pattern-based approaches that specify HCI- and software-patterns in a formalized way and respect emerging standards, can facilitate and automate the software process, and lead to solutions that can easily be adapted to varying contexts and target devices. As yet, our research was focused on adequate pattern specification formalisms and a homogeneous, at least partially automated model-driven development process topped with pattern-oriented mechanisms. However, the compilation and maintenance of the involved patterns is still carried out using a prototypical pattern editor which does not support its users sufficiently and requires time-consuming and error-prone manual intervention. Hence, we decided to replace this editor by a comprehensive pattern tool supporting the entire pattern life-cycle. Amongst other activities, we conducted a literature review of existing pattern tools during the requirements elicitation phase. The current paper describes the review approach and the considered pattern tools. Further it summarizes the related review results and our conclusions regarding the pattern tool to be designed and developed.

KEYWORDS

User interface development, model-driven development, pattern-oriented development, HCI-patterns, pattern tool.

1. INTRODUCTION

In the scope of our research we develop an integrated approach for the design and semi-automated generation of user interfaces of interactive software applications. It combines both, model-based and pattern-based development techniques and methods. The resulting framework is designed according to the *CAMELEON Reference Framework* (Calvary et al., 2002) and enables user interface designers who do not necessarily have profound software development skills to specify the diverse models which allow for at least semi-automated generation of user interface source code. The current version supports task, dialog, interaction, and layout models as well as user, device, and environment models. The complexity of the model definitions is reduced by the application of patterns of various types and different abstraction levels (Engel and Märtin, 2009). These patterns - for the most part HCI design patterns – are specified by means of a specific specification markup language which is a further refinement of the *Pattern Language Markup Language* (PLML) (Fincher and Finlay, 2003). They are stored in the framework’s core component named *Pattern Repository*. In order to significantly contribute to the model-driven treatment, patterns are equipped with model fragments which serve as building blocks for the various input models during the process of pattern application.

So far, our research effort was mainly focused on the pattern specification formalism and the pattern-based amplification of the model-driven transformation and generation processes. Thus, the patterns of the diverse types are specified by means of a merely rudimentary pattern editor. Some parts still require manual interaction, such as integrating model pieces and defining pattern interrelationships, which is an arduous and error-prone activity. To improve the framework in this regard, we decided to extend it by a tool for seamless pattern management and handling.

In this context we conducted a literature review of existing pattern tools in order to generate ideas for potential functionalities and compile a list of functional and non-functional requirements.

The rest of this paper is organized as follows: the review approach is described in Section 2, brief descriptions of the considered pattern tools are provided in Section 3, and the review results are summarized in tabular format in Section 4. Finally a discussion of the review results and our conclusions regarding the pattern tool to be designed and implemented are depicted in Section 5.

2. REVIEW APPROACH

Several pattern tool reviews have already been carried out and the results are available through the Internet, e.g. (Deng et al., 2005) and (Kruschitz and Hitz, 2009). However, these documents are fairly old and therefore do not cover novel approaches. Nevertheless, they delivered valuable input for our updated evaluation, notably for defining the evaluation criteria.

The literature review compassed the following pattern tools: UPADE (Li, 2001), (Ansari, 2003), Damask (Lin and Landay, 2002), (Lin and Landay, 2008), MOUDIL (Gaffar et al., 2003), (Gaffar et al., 2005), (Gaffar and Seffah, 2005), MUIP (Deng et al., 2005), (Deng et al., 2006), PLAss (Roski, 2006) (Engel et al., 2010), HCI PLMT (Badr and Hosny, 2006), PIM Tool (Radeke et a., 2006), Task Pattern Wizard (Seffah and Gaffar, 2007), PatternWiki (Weyers et a., 2014). Actually, also some Internet websites relating to popular pattern collections were covered by the review, including Designing Interfaces (Tidwell, 2014) (Tidwell, 2015), Patterns in Interaction Design (Welie, 2008), Design of Sites (van Duyne et al., 2009), Yahoo Design Pattern Library (Yahoo, 2014), and Quince UX Patterns Explorer (Infragistics, 2014). Due to space limits and the fact that these tools do not provide significant new functionalities far in excess of the ones listed above, they are not considered in detail within the document at hand. An overview of the considered pattern tools and their temporal grading are illustrated in Figure 1.

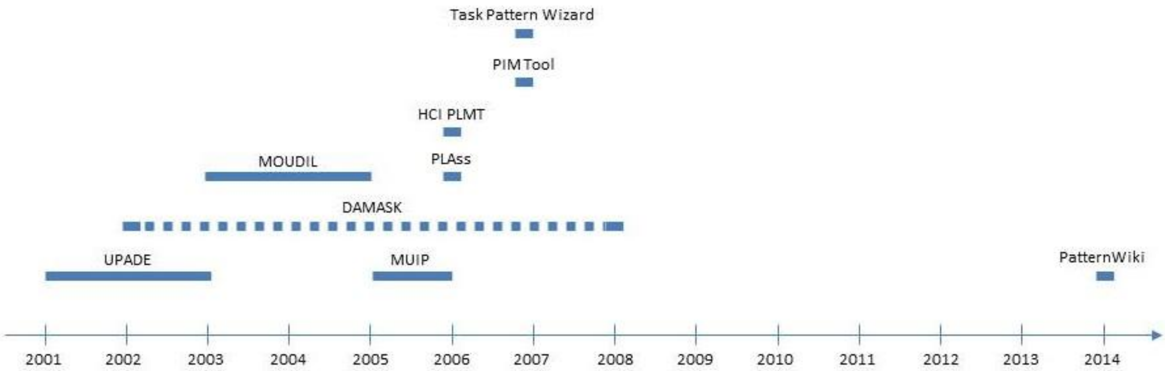


Figure 1. Overview of pattern tools considered within the literature review

Based on their respective functionality pattern tools can be divided into the three categories of pattern catalogue tools, pattern management tools, and pattern-based design tools. Catalogue tools present patterns organized in particular categories to their users. They aim at extensive dissemination of design knowhow. Pattern management tools focus on fabrication and management of patterns and pattern languages based on a unified description format. They support modification and adaptation of existing pattern specifications including inter-pattern relationships. Patterns design tools provide some pattern management capabilities but concentrate on the application of patterns within the context of pattern-based design processes (Deng et al., 2005). The reviewed tools were classified accordingly.

For each of the pattern tools we captured (1) the short name, (2) the full name, if any, (3) its originator, (4) date of first publication resp. start of development, (5) actuality in terms of current version or most current publication, (6) type of tool (pattern catalogue tool, pattern management tool, or pattern-based design tool) according to (Deng et al., 2005), (7) supported pattern description language resp. description attributes, (8) supported pattern management features (authoring, browsing, searching, modifying, relating patterns, modifying pattern collections as well as import and export of pattern descriptions) according to (Deng et al., 2005) (9) supported collaboration features, (10) supported pattern application features, (11) fundamental concepts, (12) availability in terms of the existence of a real implementation and whether it is publicly available, (13) number and type of available documentation.

Due to space limits it was not possible to present all details of the evaluated characteristics within this paper. Therefore, we decided to focus on the most important information at this point.

3. DESCRIPTION OF CONSIDERED PATTERN TOOLS

The *Usability Pattern-Assisted Design Environment* (UPADE) was developed at the *Concordia University* in Montreal (Canada) in the year 2001. Besides pattern management functions the tool supports automatic pattern application in the context of user interface design and source code generation. It is generally differentiated between pattern, design, and code abstraction levels. The patterns are specified according to the *Pattern and Component Markup Language* (PCML) (ObjectVenture, 2002). The tool provides functions to create, browse, search, modify, and relate patterns. Through the use of *Cascading Stlye Sheets* (CSS) and *Extensible Stylesheet Language* (XSL) it is possible to realize transformations for diverse target platforms, including *Extensible Hyper Text Markup Language* (XHTML), *Scalable Vector Graphics* (SVG), *Wireless Markup Language* (WML) and *Voice Extensible Markup Language* (VoiceXML). Moreover, UPADE includes a graphical editor for user interface design on the basis of pattern composition.

The pattern tool *Damask was designed at the University of California* in Berkeley (USA) in 2002. The actual implementation apparently took place as a joint work of the *IBM Almaden Research Center* in San José (USA) and the *University of Washington* in Seattle (USA). The goal of *Damask* is to create sketch-based UI designs for different kinds of user devices simultaneously. Various interaction types are supported, including Web-based styles for desktop PC und mobile phones as well as voice-based user interfaces in a prompt-and-response style. Designers create UI sketches by means of a digital pencil tool. For this, different UI elements are available, such as pages, controls, and arrows whereat pages represent Webpages or screens, controls the actual content of the pages, and arrows the relationships between the pages. Voice interfaces are modeled via prompts, responses, and forms. The latter allow for grouping prompts and responses. The UI designs are carried out in a layered approach where a general layer comprises all UI elements being available on all devices. In addition, there is one individual layer per device type containing elements which shall be available only on these particular devices. Patterns are used in order to achieve both, simplification and acceleration of building UI designs and improving UI consistency across system platforms. Patterns are specified according to the *Pattern Language Markup Language* (PLML) (Fincher and Finlay, 2003). The graphical fragments required for assembling UI designs are stored within the *solution* pattern attribute. If a pattern is selected and applied the respective graphical elements are automatically inserted simultaneously into the device-specific design layers whereat manual amendments are possible. The actual realization of the pattern is finally stored back to the *example* attribute of the pattern specification.

Like UPADE, the *Montreal Online Usability Patterns Digital Library* (MOUDIL) was developed at the *Concordia University* in Montreal (Canada) and was published in the year 2003. Target is the provision of a comprehensive framework for disseminating and interchanging design knowhow in the shape of patterns as well as their automatic application. For this, a uniform description format and a global repository for patterns are fundamentally required. The handling of the respective design knowhow follows a process called *Seven C's Methodology*: (1) “Collect” distributed information and store it centrally, (2) “Cleanup” in terms of information translation into unified representations, (3) “Certify” in the sense of the definition of a clear terminology and mapping of the acquired information to application domains, (4) “Contribute” in terms of the permanent collaborative refinement of the information, (5) “Categorize” in the sense of the definition of a uniform and well-accepted category hierarchy to which the patterns are mapped for better information handling, (6) “Connect” in terms of recognizing and establishing relations between the patterns, (7) “Control” in terms of a machine-readable format required for automatic processing. MOUDIL is intended to provide tool support for each of the seven process steps. The patterns are specified in a XML-compliant format which is not explained in detail. The framework distinguishes between the roles of pattern author and pattern user. Additionally, there is the idea to establish an international editorial board for reviewing and validating potential patterns before being published.

The pattern tool *Management of User Interface Patterns* (MUIP) was developed at the *Massey University* in Palmerston North (New Zealand) in 2005. The goal is to support users creating and maintaining patterns as well as easily integrating publicly available design knowledge. In this sense, the tool offers functions for authoring, browsing, searching, modifying, and relating patterns. It is possible to customize patterns and incorporate them into private pattern collections. In addition, pattern descriptions can be imported, exported, and translated into various description formats. The patterns are internally represented according to the *Pattern Language Markup Language* (PLML) version 1.2 which has been developed specifically for MUIP. Particular attention is paid to the pattern description attribute *forces*. It is used to describe all relevant constraints upon the pattern and the trade-off relationships between them. Further *forces* play a major role in the process of finding and selecting appropriate patterns in the context of UI design.

The *Pattern Language Assistant* (PLAss) was developed at the *Augsburg University of Applied Sciences* (Germany) in the year 2006. The tool supports two different operation modes. The open mode offers random access to both, the tool configuration parameters and the pattern resp. pattern language definitions. The other mode is read-only and allows users solely consume the pattern information. In practice the two modes realize the two discrete roles of a pattern (language) author and user. The patterns are not specified on the basis of a predetermined description format or language, but it is possible to define the description attributes individually per pattern language. PLAss provides browsing and searching functions. In open mode users can also create new patterns and modify existing ones. The specification of pattern relations is supported by an automatic module which is based on pattern hierarchy rules. The structure of the pattern specifications and the actual pattern definitions are stored in separate XML-compliant files. For this reason import and export from and to other formats can be realized by means of *Extensible Stylesheet Language Transformations* (XSLT). Out of the box PLAss facilitates transformations to HTML, PDF, and *Microsoft* DOC formats. Imports can be accomplished by manipulating the XML files outside the scope of the tool.

The *HCI Pattern Language Management Tool* (HCI PLMT) has been developed by the *American University in Cairo* (Egypt) in 2006. It is a Web application supporting its users when browsing, searching, modifying, and relating patterns. However, creating new and deleting existing patterns is not possible. Pattern specifications consist of mandatory fields for pattern name, problem, solution, Website link, and category as well as some optional attributes which are not explained in detail. The Website link incorporates a *Uniform Resource Locator* (URL) of the actual pattern definition. In order to propose potential relationships between patterns, the tool automatically evaluates the name, problem, solution, and context fields of a pattern and performs string comparisons with the specifications of the other patterns. For instance, potential alternative patterns have the same name but different Website link attributes. Supported relationship types are association, aggregation, specialization, and alternative. It is possible to change the relation type by hindsight.

The *Patterns in Modelling* (PIM) *Tool* is a joint development of the *University of Rostock* (Germany) and the *Concordia University* in Montreal (Canada) and dates back to 2007. The tool combines model-driven and pattern-based software development approaches. Target is to construct an *UI Multi Model* (UIMM) which internally comprises task, dialog, presentation, and layout models. Patterns are equipped with model segments which can be used as building blocks to extend the UIMM and therefore speed up the modeling process. In a first tool version solely task patterns are supported. The patterns are specified according to the *Task Pattern Markup Language* (TPML) (Sinnig, 2004). The application of patterns follows a four step process: (1) identification of a point within the UIMM that should be expanded by means of a pattern, (2) selection of an appropriate pattern, (3) instantiating the pattern, i.e. adapting it to the intended context of use, (4) integration of the pattern’s model segment into the UIMM. The actual model-driven generation of the user interface is not in the scope of the PIM tool and requires an external model-based software development environment.

The *Task Pattern Wizard* was developed at the *Concordia University* in Montreal (Canada) in the year 2007. It is based on the same theoretical groundwork as the PIM Tool described above. It therefore supports the selfsame and some additional models as well as the selfsame pattern application process. The creation of the models happens in three phases. In a first step the domain, task, and user models are defined. Subsequently the environment, dialog, and presentation models need to be specified. On one hand the environment model describes the physical factors impacting the user of the system to be developed. On the other hand it holds relevant information about the target user device. Within the final phase the layout model is elaborated. This approach facilitates a step by step refinement and concretion of the intended user interface until it can be automatically generated from the layout model. Patterns are used to support both, the construction of the diverse models and the required model transformation. The pattern specification attributes are not explained in detail but include at least information regarding the problem to be resolevd, the context, and the proposed solution. In addition, it is justified that the suggested solution is suitable to resolve the given problem. The model building blocks are represented by means of the *XML User Interface Language* (XUL).

PatternWiki is a joint development of the *RWTH Aachen University* (Germany) and the *University of Duisburg-Essen* (Germany) from the year 2014. The tool focuses on collaboration aspects regarding creation and maintenance of pattern descriptions. It is based on the publicly available *MediaWiki* engine which is supplemented by an extension for processing XML-compliant data formats and a graphics module facilitating the visualization of pattern relations. The patterns descriptions are specified according to the *Pattern Language Markup Language* (PLML) (Fincher and Finlay, 2003) and can be converted into other formats by means of common XML techniques, such as XSLT and *XML Schema Definition* (XSD). *PatternWiki* provides features for creating, modifying, browsing, and searching as well as for defining relations between patterns. Owing to the employment of the *MediaWiki* engine certain collaboration functions including discussion, commenting, and pattern revision are available to the users.

4. RESULTS

The most important pattern tool characteristics and features elicited during the literature review are summarized within the following Tables.

Table 2. Review results part 1.1

Pattern Tool	UPADE	Damask	MOUDIL	MUIP
Full Name	Usability Pattern-Assisted Design Environment	Damask	Montreal Online Usability Patterns Digital Library	Management of User Interface Patterns
Originator	Concordia University, Montreal	Design: Univ. of California, Berkeley Implementation: IBM Almaden Research Center, San José and Univ. of Washington, Seattle	Concordia University, Montreal	Massey University, Palmerston North
Start of Dvpt. / First Publication	2001	2002	2003	2005
Current Version / Last Publication	2003	2008	2005	2006
Availability	Existing Java prototype, executable version not located	Existing implementation, executable version not located	Existing implementation, executable version not located	Existing implementation, executable version not located
Used Pattern Description Lang.	PCML	PLML 1.1	Proprietary (not fully described)	PLML 1.2
Used Pattern Description Elements	refer to PCML	refer to PLML 1.1	<Head>, <Body>, <Relations>, <Assimilation>	refer to PLML 1.2

Table 3. Review results part 1.2

Pattern Tool	PLAss	HCI PLMT	PIM Tool	Task Pattern Wizard	PatternWiki
Full Name	Pattern Language Assistant	HCI Pattern Language Management Tool	Patterns in Modelling Tool	Task Pattern Wizard	PatternWiki
Originator	Univ. of Applied Science Augsburg	American University, Cairo	Univ. of Rostock and Concordia Univ., Montreal	Concordia University, Montreal	RWTH Aachen / Univ. Duisburg-Essen
Start of Dvpt. / First Publication	2006	2006	2007	2007	2014
Current Version / Last Publication	2006	2006	2007	2007	2014

ISBN: 978-989-8533-38-8 © 2015

Availability	Existing implementation, executable version available at Univ. of Applied Science Augsburg	Existing implementation, executable version not located	Existing implementation, executable version not located	Existing implementation, executable version not located	Existing implementation, accessible via Internet
Used Pattern Description Lang.	Any	Proprietary (not fully described)	TPML	Proprietary (not fully described)	PLML (version not specified)
Used Pattern Description Elements	Freely configurable	Mandatory: <Name>, <Problem>, <Solution>, <Web-Site-Link>, <Category> Optional: <Context>	refer to TPML	<Problem>, <Context>, <Solution>, <Rationale>	Refer to PLML

Table 4. Review results part 2.1

Pattern Tool	UPADE	Damask	MOUDIL	MUIP
Major Concepts	Three different abstraction levels: pattern, design and code level	Sketch-based cross-device UI design, prebuilt device-specific UI design fragments, which can be manually changed after being applied, created UI designs can be executed by device simulators, layer approach to distinguish general and device-specific parts	Seven C-s methodology, separated pattern author and user roles, international editorial board to review, validate, and acknowledge patterns	<forces> can be manipulated by users and support browsing, searching, and categorizing of patterns
Supported Pattern Application Features	Pattern composition, code generation	Integration of device-specific UI design fragments into UI designs	Automated context-sensitive advice for MOUDIL users, generation of Java or C# classes	none
Supported Pattern Management Features	Authoring, Browsing, Searching, Modification, Relating	Authoring, Browsing, Searching	Authoring, Browsing, Searching, Modification, Relating, Import / Export	Authoring, Browsing, Searching, Modification, Re-lating, Manipulation Compilations / Private Workspace
Supported Collab. Features	-	-	Review / Comment	Review / Comment
Tool Category	Pattern-based Design Tool	Pattern-based Design Tool	Pattern Management Tool	Pattern Management Tool

Table 5. Review results part 2.2

Pattern Tool	PLAss	HCI PLMT	PIM Tool	Task Pattern Wizard	PatternWiki
Major Concepts	Separation of author and user roles, automated support for pattern relation specification	Ontological approach, alternative and related patterns are identified on the basis of string matching operations	Patterns contain model fragments, combination of model-driven and pattern-based approaches	Patterns contain model fragments (XUL), combination of model-driven and pattern-based approaches	MediaWiki engine with extensions for XML representations and graph-based visualization
Supported Pattern Application Features	none	none	Pattern identification, selection, instantiation, and integration into UI multi-model	Pattern identification, selection, instantiation, and integration into task, dialog, presentation, and layout models	XSLT-based transformations
Supported Pattern Management Features	Authoring, Browsing, Searching, Modification, Relating, Imp. / Exp.	Browsing, Searching, Modification, Relating,	-	Browsing	Authoring, Browsing, Searching, Modification, Relating
Supported Collab. Features	-	-	-	-	Discussion, Review / Comment, Revision / Change Log
Tool Category	Pattern Management Tool	Pattern Management Tool	Pattern-based Design Tool	Pattern-based Design Tool	Pattern-based Design Tool

5. DISCUSSION

From the tool perspective, the Internet websites Designing Interfaces, Patterns in Interaction Design, Design of Sites, Yahoo Design Pattern Library, and the public version of Quince UX Patterns Explorer belong to the group of pattern catalogue tools. An exemption is Quince which is also available in versions which can be regarded as a pattern management or even design tool. MOUDIL, MUIP, PLAss, and HCI PLMT are pattern management tools, while UPADE, Damask, PIM Tool, TaskPatternWizard, and PatterWiki rank among pattern design tools.

Exhaustive pattern management support is provided by UPADE, MOUDIL, MUIP, PLAss, and PatternWiki. These tools offer capabilities for authoring, browsing, searching, modifying, and relating patterns. Additionally, MUIP provides a private workspace for individual adaptation and deployment of patterns. MOUDIL and PLAss supply their users with options for pattern import and export. Collaboration features are merely included in MOUDIL, MUIP, and PatternWiki. The former two provide features for pattern review and commenting, while PatternWiki also supports pattern discussion as well as a change log and proposing pattern revisions. In terms of formal pattern description, Damask, MUIP, and PatternWiki support the quasi-standard Pattern Language Markup Language (PLML), while UPADE utilizes the Pattern and Component Markup Language (PCML) and PIM Tool the Task Pattern Markup Language (TPML). All other reviewed tools made use of proprietary pattern description formats which are not explained in the whole.

The intended pattern tool is designed to fit into our pattern-oriented and model-driven user interface development framework and will replace the existing prototypical pattern editor. The insights gained from the current review helped us to specify and evaluate the functional and non-functional requirements regarding pattern management, user collaboration, and pattern application. In this sense, the tools and the related documentation proved to be very valuable sources of information.

In our pattern tool, the persistent storage of pattern specifications will remain unchanged, hence saved in an XML-compliant description format within the framework's *Pattern Repository*. Regarding pattern management functions we are geared to the features identified in described in (Deng et al., 2005), i.e. diverse pattern authoring, browsing, searching, modifying, and relating functions as well as manipulating forces, manipulating pattern collections, and import and export capabilities. In addition, we compass to add a user management component in order to allow registered users to create their personal workspaces where they can adapt pattern specifications to their needs and create their own pattern languages. Since we regard cooperative work and discussion on patterns as important as the related design know-how dissemination, we aim to support collaboration features as well. In this connection we plan to support comment posting, a user forum or a connatural discussion platform, and an option for proposing and maintaining pattern revisions including an adequate change log mechanism. Regarding pattern application support our framework requires the diverse model fragments incorporated within the pattern specifications. The actual model preparation and modification is not in the scope of the pattern tool itself, but is performed by means of other framework components. However, appropriate interfaces between these components and the pattern tool are required to accomplish the compulsory information exchange. The same applies to the utilization and application of the patterns within the model-driven user interface generation approach which runs outside the pattern tool.

REFERENCES

Ansari, A., 2003. *UPADE: A Tool for Automating HCI Pattern-oriented Designs*. Master Thesis, Concordia University, Montreal, Québec, Canada.

Badr, R.O. and Hosny, H.M., 2006. An HCI Pattern Language Management Tool, *Proceedings of the 2006 International Conference on Computer Engineering and Systems*. Cairo, Egypt, pp. 337-342.

Calvary, G. et al., 2002. *The CAMELEON Reference Framework*. Document D1.1 of the CAMELEON R&D Project IST-2000-30104.

Deng, J. et al., 2005. Managing UI Pattern Collections. *Proceedings of CHINZ '05*. Auckland, New Zealand, pp. 31-38.

Deng, J. et al., 2006. Focusing on a standard pattern form: the development and evaluation of MUIP. *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Compter-Human Interaction (CHINZ'06)*. Christchurch, New Zealand, pp. 83-90.

Engel, J. and Märtin, C., 2009. PaMGIS: A Framework for Pattern-based Modeling and Generation of Interactive Systems. Jacko, J. (ed.): *Human-Computer Interaction*, Part I, HCII 2009, LNCS 5610 (2009), pp. 826-835.

Engel, J. et al., 2010. Tool-support for Pattern-based Generation of User Interfaces. *In Proceedings of the 1st Int. Workshop on Pattern-Driven Engineering of Interactive Computing Systems (PEICS '10)*, Berlin, 2010, pp. 24-27.

Fincher, S. and Finlay, J., 2003. Perspectives on HCI Patterns: Concepts and Tools (Introducing PLML). *In Interfaces*, Vol. 56, pp. 26-28.

Gaffar, A. et al., 2003. MOUDIL: A Comprehensive Framework for Disseminating and Sharing HCI Patterns. *CHI 2003 workshop on HCI Patterns: Concepts and Tools*. Fort Lauderdale, USA.

Gaffar, A. et al., 2005. HCI Pattern Semantics in XML: A Pragmatic Approach. *In SIGSOFT Softw. Eng. Notes,* Vol. 30, pp. 1-7.

Gaffar, A. and Seffah, A., 2005. An XML Multi-Tier Pattern Dissemination System. *In Encyclopedia of Database Technologies and Applications,* Rivero, L. C. et al. (eds.), Idea Group.

Infragistics Inc. (2014). Quince UX Patterns Explorer. [Online] Available from: http://quince.infragistics.com/#/Main. [Accessed: 7th December 2014].

Kruschitz,C. and Hitz, M., 2009. The Anatomy of HCI Design Patterns. *Proceedings of the 2009 Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns*. Athens, Greece, pp. 202-207.

Li, N., 2001. *Usability patterns-assisted design for Web user interfaces*. Master Thesis, Concordia University, Montreal, Québec, Canada.

Lin, J. and Landay, A., 2002. Damask: A Tool for Early-Stage Design and Prototyping of Multi-Device User Interfaces. *Int. Conf. Distributed Multimedia Systems Workshop on Visual Computing*. San Francisco, USANew Zealand.

Lin, J. and Landay, A., 2008. Employing Patterns and Layers for Early-stage Design and Prototyping of Cross-Device User Interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. Florence, Italy, pp. 1313-1322.

ObjectVenture Inc. (2002). Pattern and Component Markup Language (PCML) Draft 3. [Online] Available from: http://www.cryer.co.uk/glossary/p/pcml/PCMLSpecification.pdf. [Accessed: 26th December 2014].

Radeke, F. et al., 2006. PIM Tool: Support for Pattern-Driven and Model-Based UI Development. *Proceedings of 5th International Workshop on Task Models and Diagrams for User Interface Design (TAMODIA 2006).* Hasselt, Belgium, pp. *82-96*.

Roski, A., 2006. *Entwicklung einer HCI-Pattern-Language für Websites unter Nutzung von Ergebnissen der UX-Forschung*. Dilpom-Arbeit, Fachhochschule Augsburg, Deutschland.

Seffah, A. and Gaffar, A., 2007. Model-based User Interface Engineering with Design Patterns. *In Journal of Systems and Software*, Vol. 80, Issue 8, pp. 1408-1422.

Sinnig, D., 2004. *The Complicity of Patterns and Model-based UI Development*. Master Thesis, Concordia University, Montreal, Québec, Canada.

Tidwell, J (2014). Designing Interfaces, Patterns for Effective Interaction Design (First Edition). [Online] Available from: http://designinginterfaces.com/firstedition/. [Accessed: 27th December 2014].

Tidwell, J (2015). Designing Interfaces, Patterns for Effective Interaction Design (Second Edition). [Online] Available from: http://www.designinginterfaces.com/patterns/. [Accessed: 2nd January 2015].

van Duyne, D et al. (2009). The Design of Sites, Patterns for Creating Winning Websites. [Online] Available from: http://www.designofsites.com/design-patterns/. [Accessed: 24th November 2014].

van Welie, M. (2008). Patterns in Interaction Design. [Online] Available from: http://www.welie.com. [Accessed: 27th December 2014].

Weyers, B. et al., 2014. Collaborative Creation of Interaction Patterns for the Use in User Interface Generation. *7th International Conference on Collaboration Technologies (CollabTech 2014)*. Santiago, Chile.

Yahoo! Inc. (2014). Yahoo Design Pattern Library. [Online] Available from: https://developer.yahoo.com/ypatterns/. [Accessed: 24th November 2014].

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

AUTOMATIC LAYOUT CODE GENERATION FOR ANDROID APPLICATIONS WITH MULTIPLE ORIENTATION SUPPORT

Paulo Alves[1], Porfírio Lopes[1] and Sara Paiva[1,2]
[1]*Instituto Politécnico de Viana do Castelo, Portugal*
[2]*Centro Algoritmi, Portugal*

ABSTRACT

Mobile devices have nowadays several dimensions and resolutions. We expect apps to adapt to them and present an optimized interface for each device. Sometimes, depending on the user interface, we prefer to rotate the phone so we better visualize information. However, apps have to support the orientation change and be optimized for them. This demands programmers to develop two different layouts what makes them spend twice the time. In this paper, we present a desktop application and a Android Studio plugin that aim at helping programmers to quickly generate a draft for portrait and landscape orientation based on a simple input. The plugin additionally will verify which layouts orientations are missing for a given project and automatically create them. The decision for the supported controls is based on an analysis of controls used in top used applications. In the end of this paper we present the evaluation of the solution with examples of portrait and landscape layouts automatically generated by this solution.

KEYWORDS

Automatic interface generation; Android mobile layouts; multiple orientation support; interface design rules.

1. INTRODUCTION

Android application development is continuously growing. In the fourth quarter of 2014, the Android operating system (OS) held 76,6% market share (IDC 2015) and the number of applications in the Play Store in July 2014 was of approximately 1.300.000 (Statista 2015).

A main component of almost any application, not only mobile, is the layout creation. While in other OSs such as Windows or iPhone, the Interface Development Editors (IDE) – respectively Visual Studio and XCode - support a graphical layout construction, Android Studio – the most recent IDE for Android development – does not entirely support it. Despite allowing more flexibility and better results than its predecessor Eclipse, developers often have to edit XML code directly to produce the final user interface (UI) (Hu & Zhang 2014).

Considering the wide range of Android devices with different resolutions and sizes, developers should create a layout design optimized for all of them. In addition to these aspects, the device orientation is also an important requirement for interface creation as optimizing both portrait and landscape orientation allows the application to take advantage of free space and present the user with a good UI experience. However, if creating one layout takes a given amount of time, optimizing it for landscape takes twice.

In this paper we make two contributions. The first is a desktop application that allows the definition of the interface layout in a simple way automatically creating portrait and landscape layout files in a specified location, ready to be used in Android Studio. The second contribution is a plugin for Android Studio that checks the existent layouts of a given project and generates the missing orientation layout files.

In the next section we introduce the problematic of applications with orientation support and also the structure of an android project. We refer some existent contributions and fundament the decisions made for the proposed solution. In section 3 we refer to the solution implementation, starting with an overview and then describing the desktop application, the plugin and design layout rules. Section 4 presents the evaluation of the developed solution presenting some layouts generated with it and Section 5 presents conclusions and future work.

2. PROBLEM ANALYSIS AND LITERATURE REVIEW

The importance of a custom and optimized mobile layout depending on the device orientation is clear. It allows to maximize the space usage creating a better user experience. A typical example is provided in (GoogleDeveloper 2015a) which refers to a master/detail application. In this case, in a portrait orientation in a smartphone, a list is shown and when the user clicks an item, a new activity appears with its details. If in landscape orientation or tablet, the screen can be divided in two showing the list in the left size and on the right side the details, allowing a quicker visualization. As the master/detail view is a frequent layout, Android Studio already includes a wizard that allows the creation of this behavior without additional work for the programmer. But there are innumerous other situations where the programmer has to develop two independent layouts for both orientations which significantly increases the application time delivery.

Android project folder is mainly composed of a *java* folder (for functionality implementation) and a *res* folder, which is dedicated to all project resources, such as icons, images, menus, layouts, styles, etc. Inside the *res* folder there is always a *layout* folder that holds the layout of the application and inside it there are XML (eXtensible Markup Language) files for each defined layout. If needed, the programmer can create new folders at the same level of the *layout* folder to define a layout for a specific size, density, orientation or aspect ratio. The supported sizes are *small*, *normal*, *large* and *xlarge*. The supported densities are: *ldpi* (low-density screens of ~120dpi), *mdpi* (medium density screens of ~160dpi – baseline density), *hdpi* (high-density screens of ~240dpi), *xhdpi* (extra-high-density screens of ~320dpi), *xxhdpi* (extra-extra-high-density screens of ~480dpi), *xxxhdpi* (extra-extra-extra-high-density screens of ~640dpi), *nodpi* (all densities) and *tvdpi* (screens somewhere between mdpi and hdpi of approximately 213dpi). The supported orientation are portrait and landscape. Finally, the supported aspect ratio are *long* (screens that have a significantly taller or wider aspect ratio - when in portrait or landscape orientation, respectively - than the baseline screen configuration and *notlong* (screens that have an aspect ratio that is similar to the baseline screen configuration) (GoogleDeveloper 2015b). For example, the file *res/layout/my_layout.xml* is the layout for normal screen size and the file *res/layout-xlarge-land/my_layout.xml* is the layout for extra-large in landscape orientation.

Considering the variety of options, every software that can contribute to automatically generate code is extremely important (Wen-zhen et al. 2014) and became a hot topic (Hu & Zhang 2014). One of such contributions was provided by Hu and Zhang who proposed a code generation method based on model rule to improve the development efficiency of an application and reduce the development cycle duration. The proposed solution aims to automatically generate code from a series of design rules and models, disregarding however the device orientation issue. Another relevant study is presented in (Sahami Shirazi et al. 2013) where the authors analyzed 400 apps from Google Play regarding several aspects such as multilingual and high pixel density support and also the type of used layout and controls. 67% of the analyzed apps used the LinearLayout container followed by the RelativeLayout with 24%. Regarding controls, the most used ones are the TextView (36%), ImageView (16%) and the Button (9%).

We analyzed current top applications such as Instagram, Pinterest, 9gag, LinkedIn and Facebook and concluded their layout could be created, in its majority, using LinearLayout containers and TextView, ImageView, EditText and Button controls. With this knowledge, we defined as our purpose to build a solution that helps programmers to create a first version of the layout of an application, for portrait and landscape, saving them time.

3. IMPLEMENTATION

As afore mentioned, this paper has two contributions: a desktop application and an Android Studio plugin. Both were built using the JAVA language and use a set of design rules for generating the final XML layout files. The choice for this language is due to the fact that it is also used in Android Studio and it is directly and easily added or imported into a project. In the next sections we describe the desktop application, the plugin and the layout rules.

3.1 Overview

As defined above, we use the LinearLayout container for the layouts creation. At this stage, we only require the user to specify the controls he wants and how they should be grouped. The portrait version is pretty similar to what the user specifies as he built the input in a portrait orientation too. The landscape version is optimized in terms of space and the groups are split in two columns.

3.2 Desktop Application

The desktop application´s main goal is to provide the user with a simple interface to specify the controls he wants in the interface and generate both portrait and landscape version with a button click.

The application, shown in Figure 1, first starts by asking the user the path for the *res* directory of the project he wants the generated layouts to be inserted in. This allows the user to create layouts for a project he is working on in Android Studio and as soon as the layouts are created they will automatically appear in the Android Studio structure, in the correct directories. The main interface of the desktop application (right side if Figure 1) shows the supported controls in the upper right corner of the interface which currently are ImageViews, TextViews, EditTexts and Buttons (new ones can be added with minimum effort). In the left area the user types the code of the control he wants to insert (I for ImageView and so on) and insert a line break when he wants that same effect on the portrait interface. In the example provided below the user wants an interface that has an image (the icon launcher is used as default), below the image two TextViews (side by side), below the texts an EditText and finally three Buttons (all in the same line).

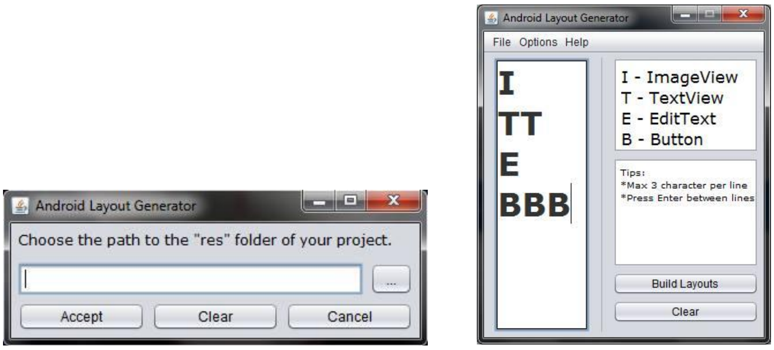


Figure 1. Desktop application for layout generation main interface

After specifying the interface controls, the user should press the “Build Layout” button that performs two operations: 1) create an intermediary file based on the controls the user inserted and 2) create portrait and landscape XML layouts based on the intermediary file.

3.2.1 Intermediary File

The intermediary file is created based on the controls the user specified for the interface and has the XML Schema Definition (XSD), represented on Table 1. The intermediary file has a XML Element named *widget* for each control and inside it two other XML elements: one that specifies the type of control (ImageView, TextView, EditText or Button) and other that specifies the group. The group value will be zero for controls that do not share the horizontal line with other controls. When a line has several controls they will all share the same group id. Different lines will have different group ids.

Table 1. XSD of the intermediary file

```
<xs:schema attributeFormDefault=”unqualified”
elementFormDefault=”qualified”
xmlns:xs=”http://www.w3.org/2001/XMLSchema”>
  <xs:element name=”Layout”>
    <xs:complexType>
      <xs:sequence>
        <xs:element name=”Widget” maxOccurs=”unbounded” minOccurs=”0”>
          <xs:complexType>
            <xs:sequence>
              <xs:element type=”xs:string” name=”Type”/>
              <xs:element type=”xs:byte” name=”Group”/>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

The algorithm followed to create the intermediary file is represented in Figure 2. The first step after reading input from the textbox is to validate the user input that mainly consists on checking if any unsupported control was inserted or if a line has more than three controls (not allowed for now). If the input is valid, each line is iterated and a different sub-process is applied for lines with one single control and for lines with more than one control. If only one control exists, the control code is transformed to uppercase (only for standardization) and after identifying the control (TextView, ImageView, EditText or Button), a new Widget tag is added to the intermediary XML file with tag *Type* equals to the control and the tag *Group* equals to 0 (which means there is no other control in that line). If the line has more than one control, the line is split and for each character a new Widget tag is added to the intermediary XML file with tag *Type* equals to the control and the tag *Group* equals to 1 (which means there is another control in the line).

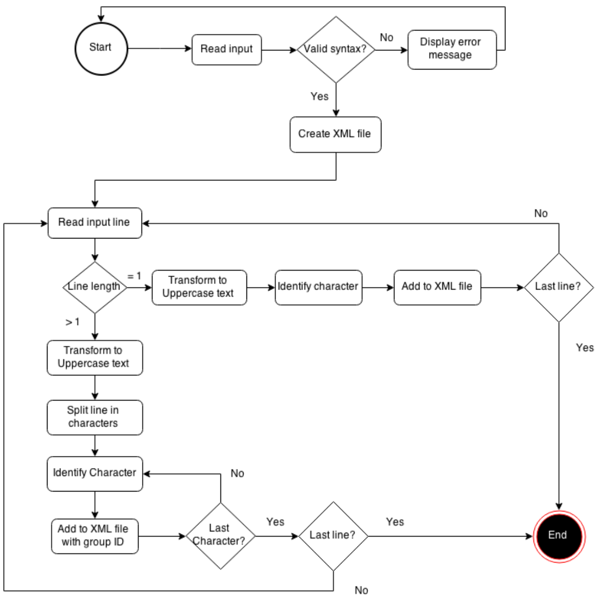


Figure 2. Flowchart for the intermediary XML file creation from the user input

3.2.2 Layouts Creation

After the intermediary file is created, a set of design rules is executed to transform the intermediary file in Android layout files (rules are explained in Section 3.4). The layout files will be created in the chosen *res* folder and Android Studio is automatically refreshed so the user can continue to edit the layout file, namely specifying the images source, the text style, margins, alignment, among others.

3.3 Plugin

The developed Android plugin has a slightly different purpose from the desktop application. The Android plugin is supposed to help programmers that have already defined a portrait version of the layout (graphically or using XML inside Android Studio) to automatically generate its landscape version or vice-versa. With a simple click in the plugin button, three situations can happen.

If by any chance there are no layout files at all, the desktop application is launched allowing the creation of a new interface that will create the correspondent portrait and landscape layouts, including their directories if needed. Android Studio automatically recognizes them and the programmer can continue to edit the layouts. If a *layout* directory exists, all layout files in it are processed and the existence of the landscape version of each one its verified. If it doesn´t exist it is created. To create a landscape layout from a portrait layout, an intermediary file (following the XSD represented in Table 1) is first created based on the portrait XML and then a landscape layout version is generated using the same layout design rules used by the desktop application and that are described in Section 3.4.

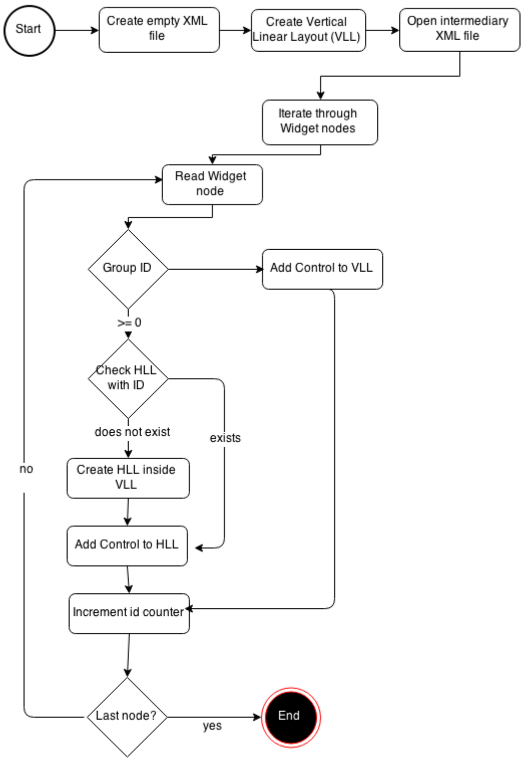
Finally, if a *layout_land* directory exists, the same process described prior for the *layout* folder is followed, assuring in the end there is a portrait version for all landscape layouts.

3.4 Layout Design Rules

In order to support the desktop application and the Android Studio plugin, two distinct set of design rules were defined: one to create portrait layout from the intermediary file (represented in the flowchart on the left in Figure 3) and another to create the landscape layout from the same intermediary file (represented in the flowchart on the right in Figure 3).

The portrait generation flowchart starts by creating a XML file and inserting a linear layout with vertical orientation (VLL). The intermediary file is opened and an iteration is made through all existent widgets. If the group id tag of the widget is zero it means it is the only control in that line so the control identified by the type tag is added directly to VLL. We increment the id counter so each control added to the XML layout file has a different identifier, which is required by Android. If the group id of the widget is greater than zero, it means the control will be inside a horizontal linear layout (HLL) with the id equal to the group id of the widget. So we start by verifying if that HLL exists (it will exist to all widget of the group except to the first one). If it does, we simply add the control to it. If it doesn't we first create the HLL inside the VLL, and then add the control to the HLL. In both cases, we also increment the id counter.

The landscape generation flowchart starts by creating a XML file and a horizontal layout with two vertical layouts inside. The purpose is to use all the horizontal space and distribute the controls through both vertical layouts. In order to decide if we put the group in the left or right layout, we start by counting the number of groups (*ng*) in the intermediary file. For each widget read, we verify if it is part of a group or not. If it isn´t, we verify the number of this group and compare to *ng*. If it is equal or less than *ng/2*, the control is added to the left vertical layout; otherwise it is added to the right vertical layout. The purpose of this reasoning is to have the same number of groups in the right and in the left if the number of groups is pair. If it is odd the left vertical layout has one more group than the right vertical layout. After adding the control to the appropriate vertical layout we increment the counter and process the next widget. In case the widget read is part of a group, the reasoning is similar to the portrait layout creation described above, where we verify if the horizontal linear layout is already created before adding the control. Whenever a new horizontal layout needs to be created the same test is performed to verify if the creation is on the left or in the right side.



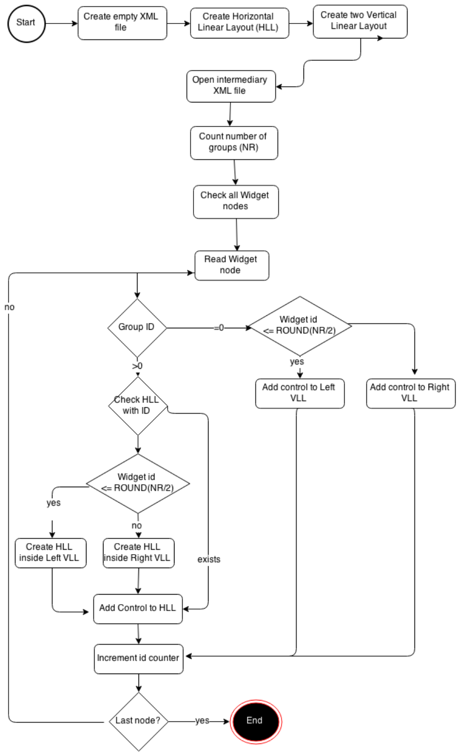


Figure 3. Flowchart for portrait (on the left) and landscape (on the right) creation

4. EVALUATION

Based on the input presented in Figure 1, the first step of the desktop application generation process creates an intermediary file with the content presented in the left side of Figure 4. The first line of the input is an image and therefore the tag *Type* of the first widget has the text *Image*. The group is zero because the image is the only control in that line. Next there are two TextView's belonging to group with id equals to one (this number is sequential). The third line of the input is an EdiText which has also group 0 in the intermediary file. Finally there are three buttons in the same line with group id equals to 2. Figure 4 also presents the portrait and landscape layouts generated from the same intermediary file, applying the design rules described above. We can notice the four vertical groups of controls in the portrait layout, where some of them have only one control (groups 1 and 3) and other have several controls (groups 2 and 4). In landscape mode, the same controls are presented differently. As there are four groups, there are two in each vertical layout. The left vertical layout has two horizontal linear layouts for the first two groups and the right vertical layout has two other horizontal linear layouts for the other groups.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<Layout>
<Widget><Type>Image</Type>
        <Group>0</Group></Widget>
<Widget><Type>TextView</Type>
        <Group>1</Group></Widget>
<Widget><Type>TextView</Type>
        <Group>1</Group></Widget>
<Widget><Type>EditText</Type>
        <Group>0</Group></Widget>
<Widget><Type>Button</Type>
        <Group>2</Group></Widget>
<Widget><Type>Button</Type>
        <Group>2</Group></Widget>
<Widget><Type>Button</Type>
        <Group>2</Group></Widget>
</Layout>
```

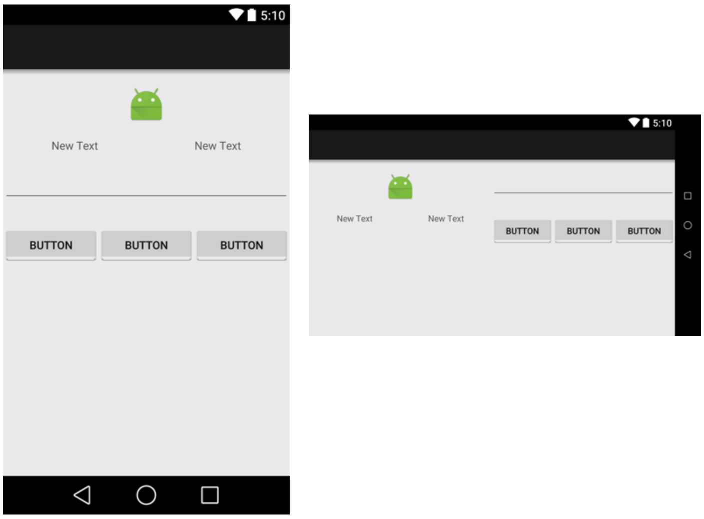


Figure 4. Intermediary file contents with the correspondent portrait and landscape layouts generated

Another generation example is shown in Figure 5. In this case we demonstrate how the tool is versatile and can be used to generate several types of layouts including those similar to well known applications such as LinkedIn or 9gag. The example shows the layout firstly created by the desktop application and after edition in Android Studio mostly replacing images and changing the text content and style.

The first layout represents a first draft built using the desktop application, simply specifying the input and with minimum effort for the programmer. The second layout shows how simply changing the images and texts a different look can be achieved, in this case to build an application similar to LinkedIn. The third and fourth layouts are an example of building layouts for an application such as QuizUp, first with the draft from the desktop application and then after editing it in Android Studio accordingly to the intended purpose.

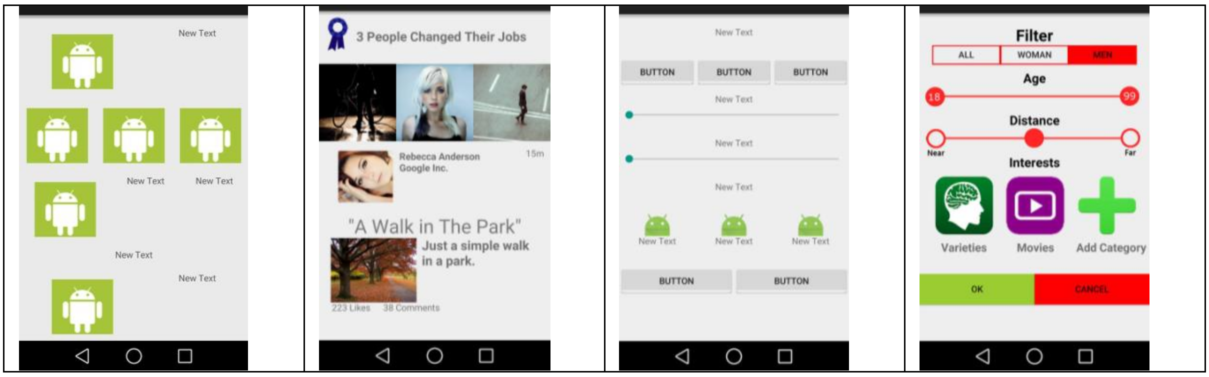


Figure 5. Creation of layout of LinkedIn and QuizUp.

Regarding the execution of the plugin, the results are pretty much the same shown for the desktop application as the design rules are the same. The layouts shown in Figure 4 can be provided as an example. When executing the plugin in an Android Studio project only with the portrait layout defined, the intermediary file representation is created with the shown content and the landscape layout created as also shown. The process is valid and similar for an Android Studio project with only the landscape layout defined.

5. FUTURE WORK AND CONCLUSIONS

In this paper we present a solution for minimizing the effort and time spent on creating Android layouts with support for multiple orientation. Creating apps with this support is an important feature as allows final users to have a better experience taking the most out of the space of the device depending on the orientation being used. As Android Studio layout creation tool is more time consuming than for example Visual Studio or XCode, software to help in the automatically generation becomes necessary.

The developed solutions include a desktop application that allows specifying the controls for the interface and with a simple button click two layouts are created, one for portrait and other for landscape. We also developed an Android Studio plugin that aims to verify the existent layouts in a given project and generate the missing ones, regarding orientation. Both solutions – desktop application and plugin – generate layout files based on design layout rules. For the portrait layout we basically use one vertical horizontal layout with several horizontal linear layouts inside where each one of these layouts have a group of controls. For the landscape layout we use a horizontal linear layout with two vertical linear layout inside splitting the groups of controls for the two vertical linear layouts so we use all the space of this orientation.

The layouts support LinearLayouts, TextViews, Buttons, EditTexts and ImageViews. We chose these containers and controls after analyzing top applications such as LinkedIn, Instagram or 9gag that are possible to build with the mentioned widgets.

Although the solution was defined specifically for the Android platform, a similar reasoning can be made for the Windows Phone OS as it also uses XML files. Regarding iOS, the mapping is not so direct as this will be left for future work.

Additionally, we will incorporate the notion of priority of a control so a user/programmer can specify the most relevant controls of the layout. This information can be important so in landscape version we do not handle each group of controls the same way, giving more emphasis to those with higher priority. We will also consider the use of relative layouts and definitely incorporate more widgets.

REFERENCES

GoogleDeveloper, 2015a. Fragments. Available at: http://developer.android.com/guide/components/fragments.html [Accessed April 28, 2015].

GoogleDeveloper, 2015b. Supporting Multiple Screens. Available at: http://developer.android.com/guide/practices/screens_support.html.

Hu, W. & Zhang, K., 2014. Research and Implementation of Android Embedded Code Generation Method based on Rule Model. International Journal of Multimedia & Ubiquitous Engineering, 9(11), pp.273–282. Available at: http://www.sersc.org/journals/IJMUE/vol9_no11_2014/27.pdf [Accessed April 28, 2015].

IDC, 2015. Smartphone OS Market Share, Q4 2014. Available at: http://www.idc.com/prodserv/smartphone-os-market-share.jsp [Accessed April 13, 2015].

Sahami Shirazi, A. et al., 2013. Insights into layout patterns of mobile user interfaces by an automatic analysis of android apps. 5th ACM SIGCHI Symposium on Engineering interactive computing systems, pp.275–284. Available at: http://dl.acm.org/citation.cfm?doid=2494603.2480308.

Statista, 2015. Number of apps available in leading app stores as of July 2014. Available at: http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [Accessed April 13, 2015].

Wen-zhen, C., Wei-feng, B. & Ya-wei, Z., 2014. Porting LCD Drive Based on Android OS. Instrumentation Technology, 14(16).

USABILITY EVALUATION ON A NEW WINDOW OPERATION INTERFACE BY FINGER GESTURE ABOVE A KEYBOARD

Junmock Lee[1], Tsutomu Terada[1,2] and Masahiko Tsukamoto[1]
[1]*Graduate School of Engineering, Kobe University, 1-1, Rokkodaicho, Nada, Kobe, Hyogo, 657-8501, Japan*
[2]*JST PRESTO*

ABSTRACT

There are many situations when users operate multiple windows simultaneously on a computer, such as looking at web browsers while making documents or searching program codes while writing programs. Generally, users can use a mouse or keyboard to change the focus among windows. However, when users frequently switch between typing and operating a mouse, they can be annoyed by having to move to shift devices, such as releasing a hand from the home position of a keyboard for mouse operation. In this paper, we propose an interface to reduce workload by enabling users to operate inactive windows when switching between typing and pointing. The interface recognizes hand gestures by using a depth sensor that images a keyboard entirely and uses the hand gestures for window operation. We implemented a prototype, evaluated task workload, and compared operation time of the proposed interface with mouse operation when typing and pointing are frequently switched between. The results suggest that using the proposed interface significantly reduces the task workload and operation time.

KEYWORDS

Input Interface, Hand Gesture, Depth Sensor, Keyboard, Window Operation.

1. INTRODUCTION

There are suitable input devices for various situations. For example, users use a keyboard for typing on a desktop computer and a mouse for pointing. Users can also use a touchpad or a pointing stick for pointing on a laptop computer. Moreover, other input devices used in unique situations such as a graphics tablet or a presentation device. Furthermore, touch screens have recently become popular because of the spread of portable devices such as smart phones and tablet computers. Also, voice recognition by a microphone and motion recognition by an accelerometer have attracted people's attention. There have been many studies on hand gesture recognition (Wachs et al., 2011): Augmented reality (Starner et al., 2000), human-robot interaction (Rogalla et al., 2002), surgeon support (Graetzel et al., 2004), home automation (Starner et al., 2000), supporting the handicapped (Liang and Ouhyoung, 1998), and so on. Because of commercial depth sensors such as Kinect and DepthSense, hand gesture recognition is recently becoming a reality.

On the other hand, a multitasking environment is provided by present operating systems such as Windows, Mac OS, or Linux, so users can submit tasks simultaneously. In a multitasking environment, users can use multiple windows, such as when they are making documents while looking at web browsers or writing e-mails while reading data. A window being used by users is called an active window, and the others are called inactive windows. When users want to activate a window, they need to click the window with a pointing device. This action is also necessary when users want to execute small window operations, such as scrolling only one line of a document. Moreover, users must do the same action to go back to a main task. Users can be annoyed by having to frequently switch between typing and operating a mouse.

The mouse is difficult to replace with the keyboard because window operations are analog and continuous. For example, by using the mouse, a user needs to move the cursor to and click only once on the taskbar to select a window to use. However, by using the keyboard, shortcut keys need to be used many times. In this paper, we propose an interface that reduces the workload of operating inactive windows when typing and pointing are frequently switched between. Using hand gestures, users can operate inactive windows without moving their hands from the home position of a keyboard or changing the focused on window. Moreover, we implemented a prototype, evaluated task workload, and compared operation time of the proposed interface with mouse operation when typing and pointing are frequently switched between.

This paper is organized as follows. Section 2 describes related works. Section 3 explains the proposed system. Section 4 details the implementation. Section 5 describes evaluation and consideration. Finally, Section 6 provides the conclusion and mentions future work.

2. RELATED WORKS

2.1 Hand Recognition and Gesture Interface

Recognizing hands and utilizing gestures for an interface by using sensors has been a research vision for decades (Rautaray and Agrawal, 2015). Ubi-Finger (Tsukada and Yasumura, 2004) controls information appliances, scrolls windows, and gives dynamic presentations by using a bending sensor, an acceleration sensor, and a touch sensor. However, when a keyboard is used, devices that are attached to hands will reduce usability of the keyboard.

There are studies that recognize hands by detecting skin color using a camera, so gestures can be recognized without additional devices attached to users' hands. The Hand Mouse (Kurata et al., 2001) points at a soft keyboard in an augmented reality (AR) environment by detecting the skin color of the user's hands by using a wearable camera. However, skin color can be affected by illumination. Especially on a keyboard, the light from a monitor changes frequently, so the screen may affect the accuracy rate of recognition. Moreover, fingers are difficult to recognize on the keyboard because they are curved and too close together.

On the other hand, Kinect, a depth sensor was developed by Microsoft. Kinect has attracted people's attention because it can detect people or things by obtaining a depth map that is not affected by illumination. OmniTouch (Harrison et al., 2001) is a wearable depth-sensing and projection system that enables interactive multitouch applications on everyday surfaces by detecting and analyzing the depth map of hands and surfaces. However, because the available range of Kinect is above 0.5 meter, Kinect requires a large space to recognize a keyboard. In spite of depth sensors that are supported in short ranges such as Leap Motion Controller by Leap Motion or DepthSense325 by SoftKinect, the above software cannot recognize hands because they are too close the keyboard to separate them.

2.2 Window Selection and Mouse Operation

When users simultaneously have many applications open on a desktop computer, they need to do many window operations. There are studies that reduce operations for selecting windows to activate some windows among several. Switchback Cursor (Yamanaka and Miyasita, 2013) operates windows that are behind the active window by making the mouse cursor go through other windows. ThumbSense (Rekimoto, 2003) clicks a mouse cursor by using the keys of a keyboard instead of a touchpad and moves the mouse cursor by using a touchpad on a laptop computer. Type-hover-swipe (Taylor et al., 2014) operates windows by using an augmented mechanical keyboard. However, the problem of switching between the active window and the inactive windows is not solved.

In this paper, we propose a system that can enable users to operate inactive windows without changing the focus or operating a mouse by recognizing hand gestures on the keyboard by using a depth sensor.

Figure 1. Setting of a depth sensor (A). Color image (B) and the depth map (C) imaged by the depth sensor.

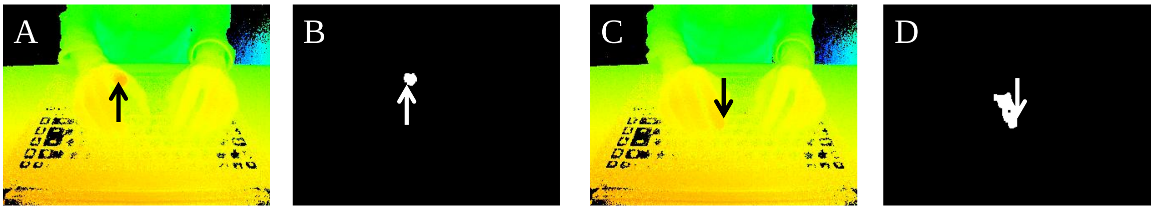


Figure 2. Depth map (A) and difference image (B) when a finger is moved up. Depth map (C) and difference image (D) when a finger is moved down.

3. SYSTEM

3.1 Finger and Gesture Recognition

Some pieces of software recognize hands by using a depth sensor: iisu by SoftKinect, Intel Perceptual Computing SDK, and so on. Current hand recognition methods typically obtain the silhouettes of hands. However, when users are typing on a keyboard, silhouettes of hands are difficult to obtain, because the hands and a keyboard are too close to separate. In this paper, we propose a method to recognize hand gestures on the keyboard. The method evaluates depth from a depth sensor and assesses a position of a finger by detecting change of a depth map caused by movements of the finger.

The depth sensor is placed on the top of a computer like in Figure 1A, thus the sensor can view a keyboard entirely. Figure 1B and C shows a color image and a depth map taken by the depth sensor. The depth sensor is taking a depth map throughout, when the application is running. The system creates a binary difference image by comparing a current depth map with a 1-frame-before depth map. A binary difference image will be saved when the depth becomes lower and the variation of a depth map exceeds the threshold. Because of the noise of the difference image, the system smoothes the image. When users make movements on the keyboard with their fingers, though users move their fingers anywhere, the depth of pixels where the finger is becomes lower. Figure 2B and D show the difference images showing pixels the depth of which decreases when users move their fingers like in Figure 2A and C. However, the system does not respond if users move their fingers on the surface of the keyboard. To obtain the point of a fingertip, the system seeks the deepest point in areas in which depth decreases. The system defines this point as the deepest point of the hand. If a user moves only one finger, the difference image shows areas of the finger, thus the deepest point must be the fingertip. If users do not move fingers, the difference image shows nothing. Therefore, the system knows when a finger is not moved and does not update the deepest point.

3.2 Window Operation

The system uses the point of the fingertip recognized by the method in the previous subsection. It is difficult for the system to point as a mouse by capturing the motion of the finger, because it cannot recognize small movement, which is required to point. It is also difficult for the system to capture the drawing gesture of the fingers, because it is difficult for users to move only one finger complexly to draw a figure. For non-continuous operation, shortcut keys are probably better than hand gestures. In this paper, we propose a system that executes operations by recognizing motions as four patterns: up, down, left, and right.

If the system only uses the finger recognition method in the previous section, the system obtains the point of the fingertip when users type normally. Moreover, to distinguish between a user's gesture and typing motion, the system uses a specific key to start gesture operation. Before a user presses the key, he/she moves a finger up. When the key is pressed, the system considers the point of the fingertip to be the standard point. If a user moves a finger up, down, left, or right while the key is pressed, the system calculates the difference between the renewed point of the fingertip and the standard point. As a result of the calculation, the system executes a window operation. For example, if a user moves a finger down while pressing the key, the system scrolls the window down. For window operations, scrolling, moving tabs, or switching windows can conceivably be done.

4. IMPLEMENTATION

Table 1. Specifications for DepthSense325

Feature	Specification
Type	Time-of-flight
Range	0.15m - 1.0m
Resolution	320 x 240 | QVGA
Noise	< 1.4cm at 1m

We implement a prototype by using DepthSense325 by SoftKinect. Table 1 shows its specifications. DepthSense325 can obtain a color image, a depth map, and sound data. However, in this paper, only the depth map is used. The system is built by C++. To analyze and visualize the depth map, OpenCV by Intel is used. To operate windows, Windows API is used.

In this implementation, the system uses the non-convert key (only for a Japanese keyboard) to recognize the start of gesture operation. The non-convert key can switch the Japanese input between *hiragana* and *katakana* alphabets. However, for a function key to also be able to switch the Japanese input, the system negates the non-convert key. The gesture recognition key in this section means the non-convert key.

In this implementation, we implement the scrolling and switching operations used for reading documents. Figure 3 shows the state transition diagram of the system and depth maps of the modes of the system. The typing mode in Figure 3 means the state when users are typing in an active window and the system is not recognizing hand gestures. If a user presses the gesture recognition key after moving a finger up, the system changes to the gesture recognition mode, and then the point of the fingertip is recorded as the standard point. If a user moves the finger up or down in the gesture recognition mode, the system changes to the scrolling mode and then scrolls a window on a second display. When a user moves a finger up and down over blue lines, the system scrolls the window up and down. Moreover, when a user moves a finger between two lines, the system does not scroll the window. When the gesture recognition key is released, the scrolling mode is over. If a user moves a finger left or right in the gesture recognition mode, the system changes to the selecting window mode and then displays three windows horizontally on the secondary display: Google Chrome, Adobe Reader, and Notepad. When the gesture recognition key is released in the window selecting mode, the window that is placed where the finger points is maximized on the secondary display. In this implementation, we choose three applications: Google Chrome, which is the most used web browser; Adobe Reader, which can read PDF files officially standardized by the International Organization for Standardization; and Notepad, which is included in Microsoft Windows. Figure 4 shows the secondary display in the window selecting mode when the user moves a finger left.

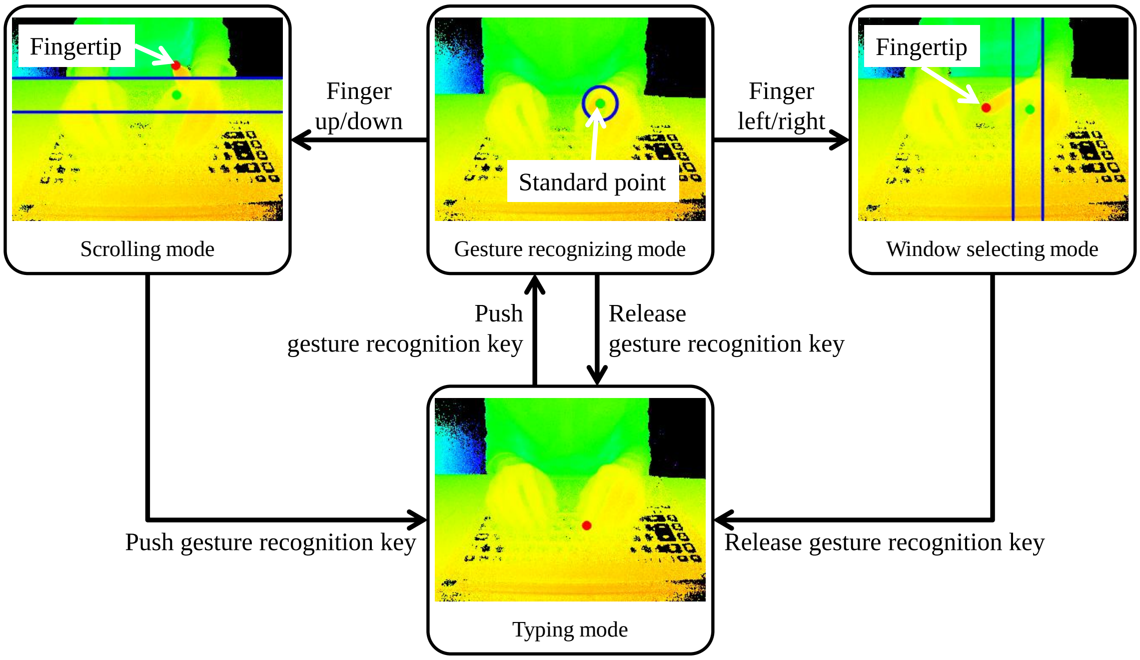


Figure 3. State transition diagram of the system and depth maps of each mode of the system.

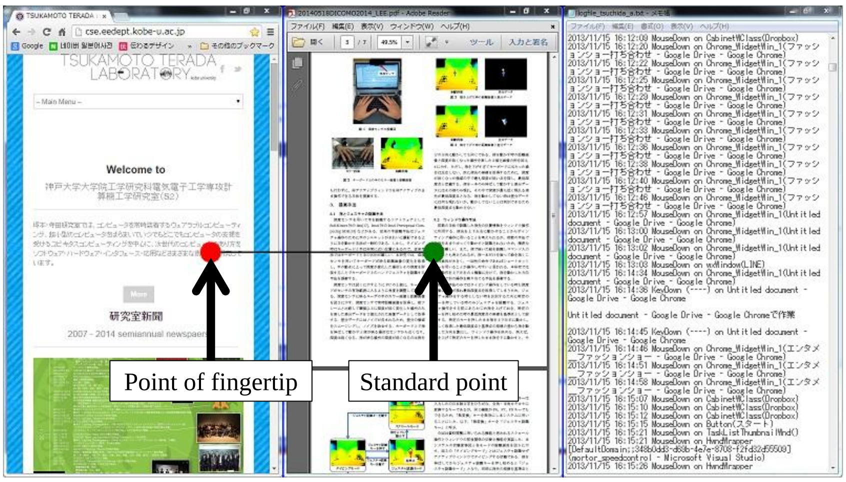


Figure 4. Secondary display in the window operation mode.

5. EVALUATION

5.1 Procedure

We evaluate task workload and operation time of the proposed system and mouse operation when a user switches between typing and pointing frequently. The user study consists of four steps. First, a subject practices using the proposed system for about 15 minutes. Second, five applications are opened on a computer: Google Chrome, Adobe Reader, Notepad, Microsoft Word, and an original key log application. About 300 pieces of Japanese text that a subject must type are written on Google Chrome, Adobe Reader, and Notepad. These three applications overlap on the secondary display. The pieces of text are randomly

ISBN: 978-989-8533-38-8 © 2015

chosen from among four paragraphs in advance. Third, the subject types the text in Microsoft Word. The subject scrolls or switches windows on the secondary display by using the system or the mouse operation. When the subject types or uses a mouse, the original application records key logs and time. The subject types the text by using the system and the mouse operation once for each. The order of interfaces is chosen randomly. Fourth, after the subject finishes typing, task workload is calculated by using NASA-TLX (NASA Task Load Index) (Hart and Staveland, 1988). We recruited six male subjects who used computers daily, ranging in age from 21 to 25, with a mean of 22. The subjects were all university students.

5.2 Results & Discussion

Table 2. Operation times

Interface	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Average
Mouse	41	41	35	47	33	42	39.8
System	28	24	39	31	29	34	30.8

Table 3. Scores of NASA-TLX

Interface	Subject A	Subject B	Subject C	Subject D	Subject E	Subject F	Average
Mouse	51.3	28.0	37.7	48.3	58.3	64.3	48.0
System	27.0	27.3	38.3	16.7	37.7	46.3	32.2

Tables 2 and 3 show the operation times of each subject and the scores of NASA-TLX, respectively. The operation time is the time for only using the proposed system or a mouse.

5.2.1 Operation Time

The average times of the mouse operation and the proposed system were 39.8 and 30.8 seconds, respectively (a 22% difference). From the results of Two-way ANOVA (Analysis of variance) for the mean of the time, the two interfaces significantly differ ($p<.05$). According to these results, the proposed system reduces time to release hands from a keyboard and operate inactive windows. Since operation and recognition mistakes waste time, it is assumed that operation time can be reduced by raising the recognition accuracy rate of the proposed system.

5.2.2 Task Workload

The average NASA-TLX scores of the mouse operation and the proposed system were 48.0 and 32.2 points, respectively (a 15.8 points difference). As a results of Two-way ANOVA for the mean of the score, the two interfaces significantly differ ($p<.05$). Thus, the proposed system is assumed to have a lower task workload than the mouse operation. Tables 4 and 5 show ratings and tallies of each subscale for each subject. The ratings are subjective scores that subjects evaluate, and the tallies are the relative importance among subscales. In this paper, the procedure of NASA-TLX is omitted.

Mental demand (MD)
There is no significant rating difference for four subjects. However, subjects C and E rated the proposed system considerably lower than the mouse operation. In our opinion, because windows selected in the window switching mode of the system are fixed in advance, it is not necessary to remember the positions of buttons on the task bar when using the system.

Physical demand (PD)
Subjects A, D, and F rated the proposed system much lower than the mouse operation. However, subject E rated of the proposed system much higher than the mouse operation. It is assumed that some users feel comfortable because they do not need to move their hands to operate the mouse while others feel uncomfortable because they need to move a finger up to have it recognized. Most subjects can move their fingers naturally. However, some subjects' fingers trembled when they moved their fingers up. It is necessary to study areas in which people can move their fingers naturally and apply the results in the system.

Table 4. Rating of NASA-TLX score

	Subject A		Subject B		Subject C		Subject D		Subject E		Subject F	
	Mouse	System	Mouse	System	Mouse	System	Mouse	System	Mouse	System	Mouse	System
MD	30	30	15	20	50	20	15	10	35	20	70	70
PD	60	10	20	10	20	15	60	15	20	50	85	20
TD	35	25	35	45	45	35	30	15	75	25	80	50
OP	40	75	20	10	20	55	40	20	70	60	15	30
EF	55	10	20	25	25	35	20	10	50	30	15	40
FR	50	25	35	25	25	20	70	20	40	30	55	25

Table 5. Tallies of NASA-TLX score

	Subject A		Subject B		Subject C		Subject D		Subject E		Subject F	
	Mouse	System	Mouse	System	Mouse	System	Mouse	System	Mouse	System	Mouse	System
MD	2	0	3	4	4	1	3	3	1	2	4	4
PD	5	5	1	1	1	1	5	2	0	0	5	0
TD	0	3	4	5	5	3	2	0	4	3	2	3
OP	1	3	2	2	1	5	1	5	3	5	3	4
EF	4	3	0	1	1	4	0	1	5	3	0	3
FR	3	1	5	2	3	1	4	4	2	2	1	1

Temporal demand (TD)
All subjects rated the proposed system lower than the mouse operation except subject B. It is assumed that the required time of the proposed system is shorter than that of the mouse operation.

Performance (OP)
Three subjects (B, D, and E) rated the proposed system lower, and the others (A, C, and F) rated the mouse operation lower. There is no consistency in the rating of performance. It is assumed that the recognition rates of hand gestures differ for different subjects, so there is a difference between subjects who can and cannot do gesture operation smoothly. Because three subjects' performance tallies were 5, which is the highest, the recognition rate must be kept high.

Effort (EF)
Three subjects (A, D, and E) rated the proposed system lower, and the others (B, C, and F) rated the mouse operation lower. There is no consistency in the rating of performance. From the results, time required to become used to the proposed system differs depending on subjects. A user study is required that evaluates time to become used to the proposed system.

Frustration (FR)
All subjects rated the proposed system lower than the mouse operation. The results suggest that all subjects could use the proposed system stably without frustration.

5.2.3 Discussion

Overall, using the proposed system reduced the average time by 22% and the average score of NASA-TLX by 15.8 points. The two interfaces significantly differ ($p<.05$). From the results, we found the usability of the proposed system to be good. Subjects used the proposed system stably without temporal demand or frustration. However, some subjects’ fingers trembled when they moved their fingers up. Thus, it is necessary to study areas in which people can move their fingers naturally. Moreover, the recognition rate must be kept high for high performance.

6. CONCLUSION

In this paper, we proposed a system that uses hand gestures to operate windows, using hand gesture recognition by a depth sensor. The system obtains a depth map for recognizing hand gestures, so users can operate inactive windows by making hand gestures. We implemented the system and evaluated task workload and operation time of the system and mouse operation when users switched between typing and pointing frequently. From the results, using the proposed system reduced the average time by 22% and the average score of NASA-TLX by 15.8 points. Moreover, the two interfaces significantly differed ($p<.05$).

For future work, it is first necessary to investigate the highest resolutions at which the system can recognize fingers. Moreover, functions must be studied that are needed for typing to implement the functions in the system. Finally, we plan to evaluate task workload of the implemented system with more subjects.

ACKNOWLEDGEMENT

This research was supported in part by a Grant in aid for Precursory Research for Embryonic Science and Technology (PRESTO) from the Japan Science and Technology Agency and by a Grant-in-Aid for Challenging Exploratory Research(25540084) from the Japanese Ministry of Education, Culture, Sports, Science and Technology.

REFERENCES

Graetzel, C. et al., 2004. A non-contact mouse for surgeon-computer interaction. *Technology and Health Care*, Vol. 12, No. 3, pp. 245-257.

Harrison, C. et al., 2011. OmniTouch: wearable multitouch interaction everywhere. *In Proceedings of the 24th annual ACM symposium on User interface software and technology*, Santa Barbara, USA, pp. 441-450.

Hart, S. G. and Staveland, L. E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, Vol. 52, pp. 139-183.

Kurata, T. et al., 2001. The hand mouse: Gmm hand-color classification and mean shift tracking. *Proceedings IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, Canada, pp. 119-124.

Liang, R. H. and Ouhyoung, M., 1998. A real-time continuous gesture recognition system for sign language. *In Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, pp. 558-567.

Rautaray, S. S. and Agrawal, A., 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, Vol. 43, No. 1, pp. 1-54.

Rekimoto, J., 2003. ThumbSense: Automatic input mode sensing for touchpad-based interactions. *In CHI'03 Extended Abstracts on Human Factors in Computing Systems*, Fort Lauderdale, USA, pp. 852-853.

Rogalla, O. et al., 2002. Using gesture and speech control for commanding a robot assistant. *In Proceedings of 11th IEEE International Workshop on Robot and Human Interactive Communication*, Berlin, Germany, pp. 454-459.

Starner, T. et al., 2000. MIND-WARPING: towards creating a compelling collaborative augmented reality game. *In Proceedings of the 5th international conference on Intelligent user interfaces*, New Orleans, USA, pp. 256-259.

Starner, T. et al., 2000. The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. *In the fourth international symposium on Wearable computers*, Atlanta, USA, pp. 87-94.

Taylor, S. et al, 2014. Type-hover-swipe in 96 bytes: A motion sensing mechanical keyboard. *In Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, Toronto, Canada, pp. 1695-1704.

Tsukada, K. and Yasumura, M., 2004. Ubi-finger: a simple gesture input device for mobile and ubiquitous environment. *Journal of Asian Information, Science and Life*, Vol. 2, No. 2, pp. 111-120.

Wachs, J. P. et al., 2011. Vision-based hand-gesture applications. *Communications of the ACM*, Vol. 54, No. 2, pp. 60-71.

Yamanaka, S. and Miyashita, H., 2013. Switchback Cursor: Mouse Cursor Operation for Overlapped Windowing. *In Human-Computer Interaction–INTERACT 2013*, Cape Town. South Africa, pp. 746-753.

USER EXPERIENCE OF DIGITAL DIORAMAS FOR INTERACTIVE WHITEBOARD

Carlo Emilio Standoli[1], Fiammetta Costa[1], Alessandra Mazzola[1], Sabrina Muschiato[1] and Daniela Zambarbieri[2]
[1]*Politecnico di Milano – Dipartimento di Design, via G. Durando, 38/A – 20158 Milano (MI)*
[2]*Università di Pavia – Dipartimento di Ingegneria Industriale e dell'Informazione, via A. Ferrata, 5 – 27100 Pavia (PV)*

ABSTRACT

This paper presents the methods and procedures applied to evaluate a Digital Diorama: a multimedia product for Interactive Whiteboards, designed to support science teaching in primary schools and specifically regarding environmental issues.
The study is part of a wider interdisciplinary research conducted through the collaboration between Academia (Politecnico di Milano, Dipartimento di Design; Università degli Studi di Milano-Bicocca, Dipartimento di Scienze Umane per la Formazione; Università di Roma Tor Vergata, Dipartimento di Scienze e Tecnologie della Formazione), primary schools and other public institutions for scientific knowledge diffusion.
Its overall purpose is to investigate whether Digital Dioramas, used with Interactive Whiteboard Technology, can improve the teaching of some fundamental aspects of biology (with special reference to the ecology), geography, and other scientific content. The applied method is innovative as it combines qualitative behavior observation with eye tracking methodology, which represents a quantitative and objective evaluation tool.
Preliminary experimental results seem to indicate that the fruition of Digital Diorama is fairly easy and it stimulates the interaction among kids and their curiosity through emotional and cognitive captivation. Furthermore, Digital Diorama can support teachers in their classroom activities, even more than the typical tools of formal learning such as readers.

KEYWORDS

User interface, eye tracking, behavior observation, usability, interactive whiteboard.

1. INTRODUCTION

Digital Diorama (DD) is an adaptive-learning educational model based on the use of computers and interactive teaching devices, able to turn the learner from passive receptor of information into collaborator in the educational process. The interest in developing digital dioramas to reproduce the experience of dioramas' exploration in museums is already stated in literature (Narumi et al., 2011).

Previous studies regarding the use of physical dioramas to elicit ecosystem complex thinking in children demonstrated the strength of affective learning (Gambini et al., 2008). Dioramas, as scientific models of natural ecosystems, lead users from virtual environments to real-life learning situations.

Dioramas, as scientific models of natural ecosystems, lead users from virtual environments to real-life learning situations; previous studies regarding the use of physical dioramas to elicit ecosystem complex thinking in children, demonstrated the strength of the induced affective learning (Gambini et al., 2008).

Their digitalization and fruition through Interactive Whiteboards (IWBs) dedicated platforms is intended to permit a wider exploitation of these resources for learning and teaching, inside and outside the classroom (museums, parks and public workstations).

Nowadays IWB device is an increasingly global educational technology and a further rapid growth is expected in the next five years. The market research company FutureSource Consulting, considering 34 million teaching spaces, found that one in eight classroom across the world now has an IWB device and, by 2015, one in five will have one (Avvisati et al, 2013). However, few multimedia products have been specifically designed for this type of technology up to now. Furthermore, usability and acceptability are crucial for their effectiveness.

Data collection and analysis regarding primary school children behavior is relevant to support the development of the interactive DD system (Nesset and Large, 2004; Andreoni et al., 2010). To this purpose, users can be observed while interacting with digital prototypes in laboratory and in the real context of use (Hanna, 1999). In order to study user experience and behavior, experiments were conducted within laboratory recreating similar conditions to the classroom environment.

An innovative method was applied combining the observation of physical behavior and the analysis of eye tracking. The use of eye tracking represents a methodology able to provide experimental evidence for objective and quantitative analysis of subjects' visual exploratory behavior in many situations of everyday life (Hayhoe and Ballard, 2005; Zambarbieri, 2003; Zambarbieri and Carniglia, 2012). In fact, clear vision of an object is guaranteed only when its image falls within the central part of the retina, which is called fovea. Therefore, in order to explore a visual scene, the eyes have to move in such a way as to bring the image of an object of interest onto the fovea. During visual exploration, the eyes make saccades and fixations. Since visual information is acquired by the central nervous system only during fixations, whereas saccades are used to shift the gaze from one point to another, it is reasonable to infer that the eye tracking methodology represents a powerful tool for the study of exploration strategies and the underlying cognitive process. Thus, when a subject is exploring a visual scene eye movements supply information about the focus of subject's attention.

The contemporary use of videotaping technologies permits to apply a behavior analysis model regarding different interaction modes and to work in interdisciplinary groups composed by engineers, designers, social scientists, usability experts. The analysis of the kid's spatial behavior, as well as psycho-perceptual and emotional expressions, is performed through a software for video annotation, to create a structured database of the different types of interaction.

The aim of this paper is to describe the methodology and the results of a preliminary experimental phase designed to verify usability of the interface and observe children's behavior. The following steps of the project will include the extension of the examined population of children; the adjustment of the interface to improve the usability of control keys and finally the experimentation of the DD within the real context of use in the classroom.

2. THE DIGITAL DIORAMA

The Tursiope DD (from Museo di Storia Naturale of Livorno) analysed in this study, represents a cut of Mediterranean Sea with its specific animals and vegetation. The DD is a digital photographic reproduction of the real diorama and it consists of a multimedia interface containing touch-screen hotspots that give the possibility to explore DD. The hotspots provide three different types of fruitions, based on an active and captivating interaction with the user: scientific in-depht analysis, link with aspects of everyday life, contents for teachers.

The DD is structured in four levels:

- Level 1, Imaginary Trip ("Viaggio Immaginario"): emotional exploration of DD through the listening of audio descriptions;
- Level 2, Crossover Topics ("Tematiche Trasversali"): scientific multimedia contents (videos, diagrams, photos, short taglines) that lead to reasoning about the relationships between individuals, other biological elements and everyday life;
- Level 3, Identity Card ("Carta di Identità"): specific information organized in FAQ and related answers about the DD elements;
- Level 4, Practical Activities ("Attività Pratiche"): suggestions for practical experiences to be unrolled at school.

These levels mix different models of intentional and non-intentional learning, taking advantage of non-formal and informal learning, (Trinder, 2008) embedding learning in planned daily activities, that are not explicitly designed as learning tools.

The development of the Tursiope DD is based on an iterative prototype evaluation process, a consolidated methodology largely applied in the field of computer interfaces (Bury, 1984) comprehending free user exploration, expert analysis and user tests which led to the development of the first release of the DD, now in use in the first test elementary school. The different phases of prototype development concerned the solution of technical problems and the improvement of functional performances required for an easy use of the DD. A relevant phase in this process are the experiments described in this paper.

3. MATERIAL AND METHODS

3.1 Experimental Protocol

Four primary school students (two girls and two boys) aged between 7 and 8 years, have been involved. This number of test users seems to be enough to detect around 70% of the existing usability problems and to achieve a 60% ratio of benefit to cost (Nielsen and Landauer, 1993) which is a result coherent with the goal of this preliminary experimental phase.

The experiments were organized in two steps:

- Step 1: each kid, wearing eye tracking glasses, stood up in front of the IWB device. The experimenter guided the diorama exploration by asking the kid to execute 29 specific tasks. The test lasted about 10 minutes per kid;
- Step 2: the four kids were together in front of the IWB device and the experimenter asked them general questions about the contents of the diorama. The test lasted about 25 minutes.

Throughout the entire experiment, the subjects were video-recorded by means of a camera placed in the upper corner of the experimental room, behind the children.

3.2 Eye Movements Recording

Eye movements are recorded by video-oculographic technique (VOG). VOG makes use of the image of the eye taken by a digital video camera to compute gaze direction. In order to make this data processing fast enough to be performed in real time, the image of the eye is illuminated with infrared light to create corneal reflexes. The infrared light is not visible to the subject so it will not cause a distraction, however it is visible to the camera.

Eye Tracking Glasses (SMI – SensoMotoric Instruments GmbH, Germany) are used in this study (Figure 1). Within the lightweight frame of the glasses six infrared light emitting diodes illuminate the eye to create the corneal reflexes. In the frame, also a digital camera is embedded to collect the eye images. A further camera, placed in the front of the glasses is applied to record the scene in front of the subject, allowing the correlation between gaze position and the environment.

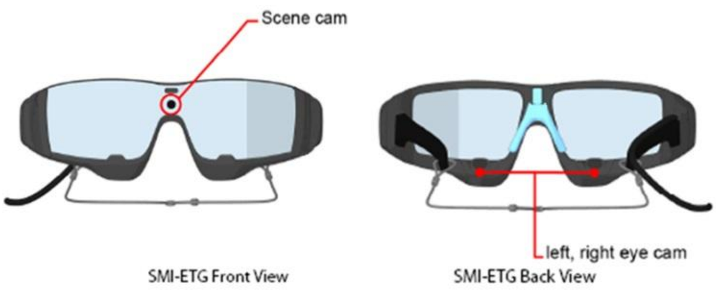


Figure 1. Front and back view of the SMI eye tracking glasses, showing cameras position

3.3 Eye Movements Analysis

Eye movement analysis was performed offline by using the BeGaze software. The first stage in the analysis consists of the identification of saccades and fixations within the recorded eye movements. To identify fixations, BeGaze makes use of a dispersion-based algorithm with a maximum dispersion threshold set at 100 pixels, and a minimum fixation duration of 80 msec.

Gaze displacement between two successive fixation positions are identified as saccades. Once saccades and fixations are recognized from the temporal sequence of eye position data, the correlation between gaze position and the scene taken from the camera is established. From the position and duration of all fixations, the scanpath is reconstructed: fixations are represented by circles whose diameter is proportional to fixation duration, whereas straight lines represent saccades. Within each screenshot of the Diorama, areas of interest (AOI) can be defined and BeGaze calculates for each AOI several eye movement parameters.

3.4 Physical Behavior Observation

The main objective of the Step 2 is to define what attitudes can be identified among the four children, in order to understand whether the DD can help them to improve network learning and their capacity to connect topics and develop content in group. The secondary aim is to observe how children behave when they use the IWB interface during test, in particular, to mark any signal of cognitive difficulty, fun behavior, boredom and curiosity during interaction.

Step 2 was organized in two parts: the first part, lasting 20 minutes, was a semi-structured trial to understand whether there was a collaboration among children during the exploration of the DD. The second part was conducted in form of a 5-minute session in which children were completely free to interact with DD. During the exploration of various content of the DD, the researcher submitted open-ended questions to the children, who were free to answer and express themselves.

Qualitative analysis was based on the direct observation performed by a behavior expert, considering: emotional involvement, attention conservation and kids' ability to create discussion on crossover topics. Quantitative analysis was performed examining video-recordings through a software for video annotation called Advene (Annotate Digital Video, Exchange on the Net) in order to identify kids' behaviors and verbalizations.

The identified activities were organized in two classes (Figure 2):

- physical participation of the individual child and the group: spatial behavior (task: move);
- verbal participation of the individual child and the group: oral interaction (task: talk).

Afterwards, the analysis grid was expanded to collect data regarding curiosity elements, stimulation of observation, group interaction and group discussion.

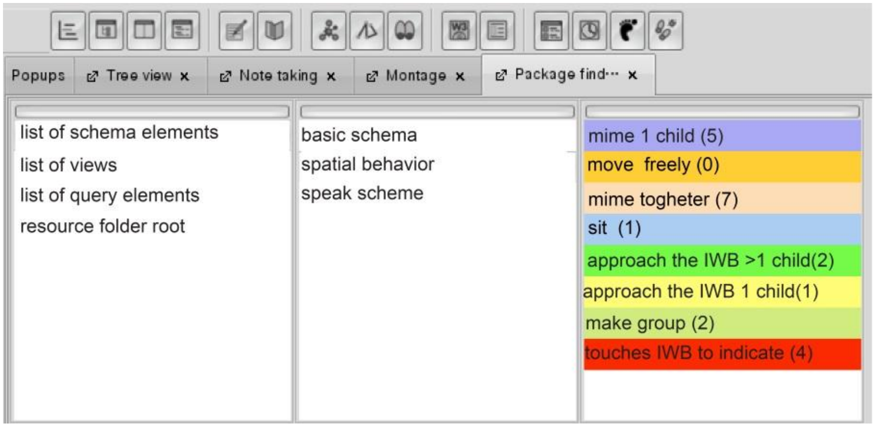


Figure 2. Schemas of verbal interaction and spatial behavior

4. RESULTS

Among the whole 29 tasks requested to the subjects in Step 1, the 11 more comparable ones were deeply analyzed and three of them were selected for detailed results presentation, as the most representative of user interaction with IWB device. Results are described both in terms of gaze analysis and behavior observation. In the last paragraph (4.4 Group test) the behavioral observation of Step 2 is reported.

4.1 Task 1

Subjects were requested to reach one level ("Viaggio Immaginario" or "Tematiche Trasversali") of the DD from the home page. To accomplish the task subjects had to localize the menu control keys placed in the bottom left of the screen, touch it to extend the menu and visualize items (Figure 3). Three subjects correctly reached the goal in few seconds without researcher's support, whereas the fourth subject needed instructions to recognize the menu.

The scanpath in Figure 3 shows an example of subject behavior during the search of "Viaggio Immaginario" level: the subject's gaze first points the requested item, and then it moves throughout all items before going back to "Viaggio Immaginario". Only at that time, the subject selects the requested option.

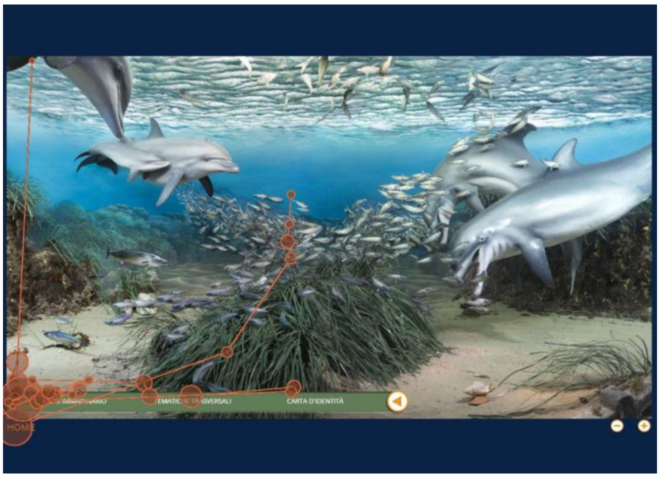


Figure 3. Scanpath of a subject during the execution of task 1. Circles represent fixations, straight lines saccades

4.2 Task 2

Subjects were requested to exit from the hotspot and to return to the level main page. Different options are available to achieve the task:
1. Pushing the zoom buttons, in the lower right corner;
2. Pushing the level title at the bottom of the page;
3. Pushing "HOME" and, from the home page, re-selecting the level.

ISBN: 978-989-8533-38-8 © 2015

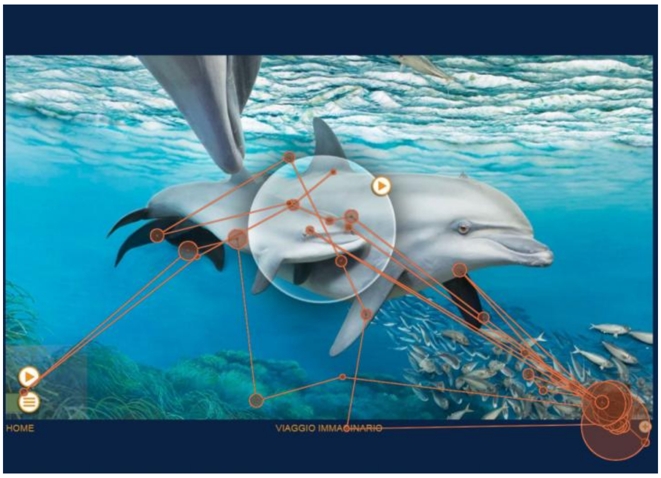


Figure 4. Scanpath of a subject during the execution of task 2

During the experiment, this task was submitted three times. The expected behavior was that the subjects chose option number 1. Actually, they did it, but at the first task's appearance, the execution was not immediate. In fact, it emerged from the scanpath analysis that subjects first explored the lower area of the screen where all control keys are placed and only later, they found the zoom buttons (Figure 4). As the task was repeated, it emerged that the subjects learnt to immediately reach the zoom buttons.

4.3 Task 3

Subjects were requested to exit the actual level and return to the previous. There are different options to achieve the task:

1. Pushing the leftward arrow;
2. Pushing the level title at the bottom of the page;
3. Pushing "HOME" and, from the home page, re-selecting the level.

During the experiment, this task was submitted two times. Three subjects correctly found the leftward arrow in the lower right corner since the first time, as shown by the Scanpath in Figure 5. The fourth subject behaved differently in both tasks presentation, by using option 3 and 2, respectively.

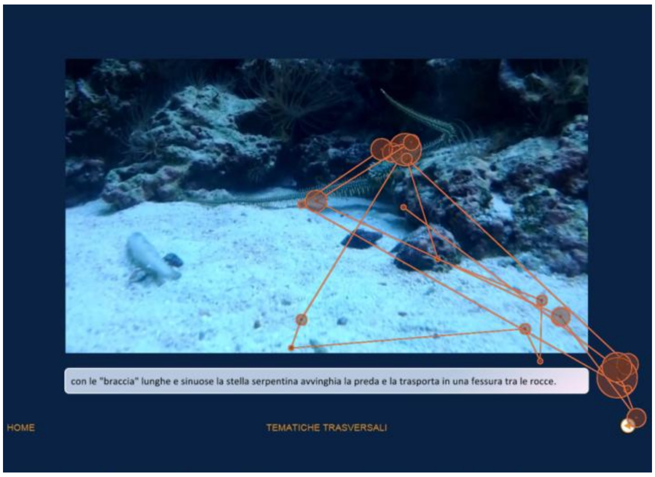


Figure 5. Scanpath of a subject during the execution of task 3

4.4 Group Test

At the beginning of the test, children were gathered and seated around the table in front of the IWB. They remained seated for one minute and then stood up all together and cooperated in the choice of the topic to explore.

In general, all children were actively involved in the test. When a child began to speak aloud or mimed animals' behavior, the mates were stimulated in enjoying the diorama also with body movements. All children coherently answered the questions and they often started spontaneous discussions on different topics.

Children participated for the whole duration of the test with a high interest. Especially during the first part, male children were more active in movement compared to females: males more often motivated others to participate in the interaction.

During the last 5 minutes of the test, an important behavioral change occurred: when the researcher moved away and left children alone, they kept on interacting with the DD. By using controls, they led the exploration of diorama and stimulated discussion on the topics, based on their free choice.

A timeline chart, structured by an expert in psycho-pedagogical sciences, permitted to detect general objectives and related tasks, identifying children verbal and non-verbal behaviors (Figure 6).

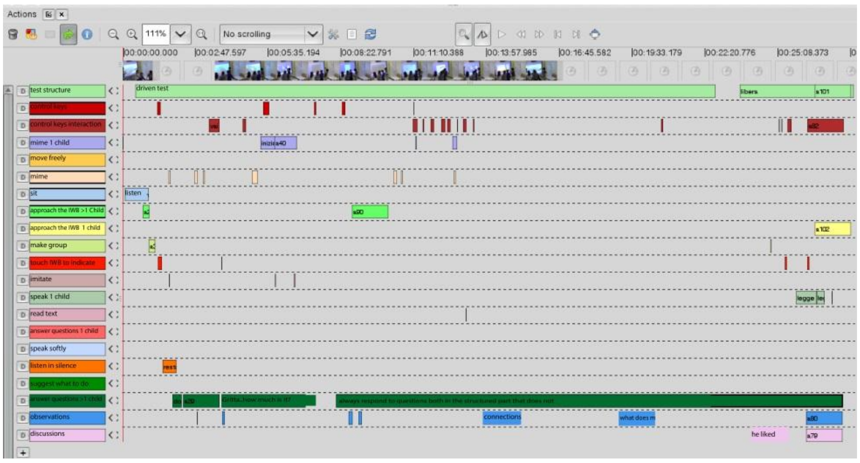


Figure 6. Example of verbal and non-verbal behavior mapping

Comparison of the four subjects shows that to mime animal movements and actions was a recurrent behavior which occurred 7 times during the trial with a total duration of almost three minutes. Another interesting spatial behavior observed was group building. In two annotations of the test, it was obvious that children got together to select or deepen a particular topic (Figure 7).

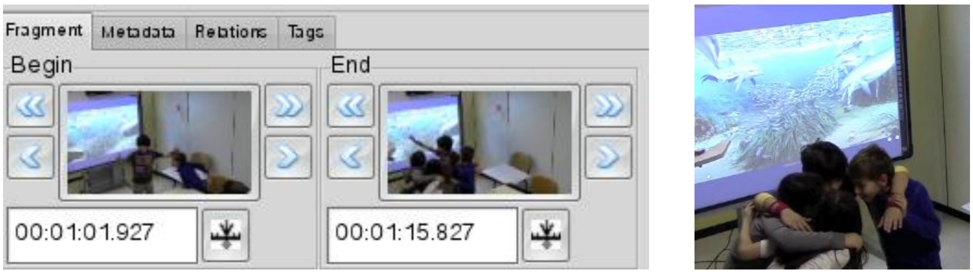


Figure 7. Example of group building and the related picture

Data collected on verbal and non-verbal behavior revealed that kids linked different elements of the DD, jumping from one topic to another, putting them into relationship, reasoning and discussing individually and in group. Several individual observations and group discussion have been recorded for a total amount of respectively 4.40 minutes and 2.50 minutes. Their remarks were ascribable to reflections on the proposed themes, displaying their perception of living systems' complexity.

5. CONCLUSIONS

Experimental evidence about user interaction with DD were obtained by using qualitative and quantitative methodological approaches. Quantitative approach was based on the recording and the analysis of eye movements during the exploration of the visual scene; qualitative approach is based on video-recording analysis for the identification of verbal and non-verbal behaviors.

Data described in this paper represent the result of the preliminary experimental phase of a multidisciplinary research project. Results seem to suggest that kids learn very quickly how to interact with the DD. In fact, children are able to complete tasks execution by locating the correct element within the interfaces as shown by the scanpath of the eyes. In some circumstance they do not identify the requested item (menu key, level title) immediately although they search in the right place, and need suggestion from the researcher to recognize it. This visual behavior suggests that the location is appropriate but their redesign could help to facilitate identification and avoid misleading messages. Another proposal to simplify interaction could be the presence of a back key in every page.

Regarding behavior observation, research findings show that the fruition of digital diorama stimulates the interaction between kids and their curiosity, through emotional and cognitive captivation and otherwise can support teachers in classroom activities even more than the typical tools of formal learning. The engagement produced by video contents suggests to replace static images with videoclips in the hotspots to support group discussion and reasoning. A difference in male and female engagement has also been observed.

The evaluation of this finding together with the validation of possible adjustments in the interface is planned in the following steps of the project involving more children and experimenting the DD in the real context of use.

Recent results in the literature have examined reading activities and comprehension on IWBs (Shams and Dabaghi, 2014; Megalakaki et al., 2015) and the learning effects (Şad and Özhan, 2012). Our results, focused on the user experience with DD provide further evidence about usability and effectiveness of IWBs underlining the potentialities offered by IWB device in didactic contests.

To spread the diffusion of the DD as an innovative learning resource, alternative IWB open source hardware technology can be designed and realized to lower the cost of installation in schools. and spread the diffusion of “Digital Diorama”. This will provide a valuable tool for the understanding of many ecosystems and the most important basic concepts of ecology, a method using guided participation and collaborative learning in a framework to support responsible attitudes and sustainability practices in everyday life. The DD can also be deployed on the web or thanks to workstation located in public places such as urban parks, municipalities, museums, since it is important that the learning of new concepts continues beyond school, throughout life to make each individual a true member of the society of knowledge.

ACKNOWLEDGEMENT

Work supported by MIUR (*Ministero dell’Istruzione dell’Università e della Ricerca*): within *Bando per la diffusione della Cultura Scientifica 2012*.

REFERENCES

Andreoni, G. et al, 2010. Sustainable Children's Product Experience. *Proceedings of Sustainability in Design: Now!* Bangalore, India, pp. 1184-1195.

Avvisati, F., et al., 2013. Review of the Italian Strategy for Digital School. *Center of educational research and innovation.* OECD Publishing.

Bury, K.F., 1984. The iterative development of usable computer interfaces. *Proceedings of IFIP INTERACT'84 International Conference on Human-Computer Interaction*, London, U.K.

Gambini, A. et al., 2008. Progetto diorami. Un percorso didattico costruito sulla collaborazione tra Università e Museo. *Atti del VI Convegno Nazionale sulla Comunicazione della Scienza*. Polimetrica Publisher. Forlì, Italy, pp. 153-160.

Hanna, L. et al. 1999. *The role of usability research in designing children’s computer products*. In A. Druin (Ed.), *The design of children’s technology*. Kaufmann, San Francisco:, pp. 4-26.

Hayhoe, M. and Ballard, D., 2005. Eye movements in natural behavior, *Trends Cognitive Scences*. i 9, pp. 188-194.

Megalakaki, O. et al., 2015. Assessing visibility, legibility and comprehension for interactive whiteboards (IWBs) vs. computers. *Educational Psychology*, ahead-of-print, pp. 1 – 20.

Narumi, T. et al., 2011. *Digital Diorama: AR Exhibition System to Convey Background Information for Museums*. In R. Shumaker (Ed.), *Virtual and Mixed Reality*, Part I, HCII 2011, LNCS 6773, Springer-Verlag Berlin Heidelberg, Germany, pp. 76–86.

Nesset, V. and Large A., 2004. Children in the information technology design process: A review of theories and their applications. *Library & Information Science Research*, n 26, pp. 140–161.

Nielsen, J., and Landauer T. K., 1993. *A mathematical model of the finding of usability problems*. Proceedings ACM INTERCHI'93 Conference. Amsterdam, the Netherlands, pp. 206-213.

Şad, S., N. and Özhan, U., 2012. Honeymoon with IWBs: a qualitative insight in primary students' views on instruction with interactive whiteboard. *Computers & Education*, 59 (2012), pp. 1184-1191.

Shams, N. and Dabaghi, A., 2014. Iranian Learners L2 Reading Comprehension: The Effect of Online Annotations via Interactive Whiteboards. Journal of English Language Teaching and Learning, N° 14, pp. 37 – 56.

Sterling, S. L., 2001. *Sustainable education: re-visioning learning and change*. Green Books for the Schumacher Society, London, U.K.

Trinder, K. et al., 2008, Learning from digital natives: bridging formal and informal learning, *The Higher Education Academy Final Report*.

Zambarbieri, D. 2003, E-TRACKING; Eye tracking analysis in the evaluation of e-learning systems, *Proceedings HCI International Conference 2003*, Vol. III, pp. 617-621.

Zambarbieri, D. and Carniglia, E., 2012. Eye movement analysis of reading from computer displays, eReaders and printed books. *Ophtalmic Physiol Opt*, 32, pp. 390-396.

ISBN: 978-989-8533-38-8 © 2015

A COMPARISON OF IMMERSION BETWEEN PLAYERS PLAYING THE SAME GAME WITH AND WITHOUT GRAPHICS

Henrik Engström, Jenny Brusk and Per Anders Östblad
University of Skövde, School of Informatics, Box 408, 541 28 Skövde, Sweden

ABSTRACT

The aim of the project presented in this paper is that visually impaired and sighted players should be able to play the same game and share a gaming experience. The goal is that the game should be accessible to visually impaired players without any additional tools, such as text-to-speech, that may reduce the immersion. At the same time, sighted players should perceive the game as a regular game. This paper presents an evaluation of the game where the player immersion has been evaluated through a post test immersion questionnaire. The study was conducted with three independent groups: sighted players using graphics (n=10), blindfolded sighted players (n=10) and visually impaired players (n=9). Although progress in the game and the reported sense of control differed between groups, player immersion was very high in all groups. There were differences between the three groups only in one out of five immersion factors. The result shows that it has been possible to provide an immersive experience irrespective of whether the players are playing the game with graphics or using audio only.

KEYWORDS

Inclusive game design, immersion, questionnaire, audio games

1. INTRODUCTION

Although digital games have a broader scope than video games, there is still a very strong emphasis on the visual dimensions, which makes most parts of the game culture inaccessible to visually impaired persons.

There exist a few audio-only games particularly targeting visually impaired players and games that they can play by the assistance of various aids, such as the VoiceOver available on iOS devices. However, there are hardly any examples of graphical games that in their original design allow visually impaired and sighted players to play on equal terms and thereby making it possible for them to share the same gaming experience. One could argue that there is a difference between making a game accessible and making a game inclusive. Accessible means it is possible to play, inclusive means that the group to include is considered in every aspect of the game development process and that the gaming experience is as similar as possible regardless of whether the player have access to the graphics or not.

This paper presents a mobile game developed with the aim to be inclusive for both visually impaired and sighted players. The goal is that they should be able to play the same game and have a very similar experience. The game has been evaluated in a study where the self-reported immersion from players in three different groups are compared. The two main groups are visually impaired subjects and sighted subjects playing with graphics. The third group, sighted subjects playing without graphics (blindfolded), forms a control group. The instrument used in this study is to a large extent based on the *Immersive Experience Questionnaire* (IEQ) developed by Jennett et al. (2008). The IEQ was specifically made for studies of computer game immersion. In this study, we have adjusted IEQ to be usable for studies that includes visually impaired players.

2. BACKGROUND

2.1 Inclusive Game Design and Audio Based Games

According to Barlet & Spohn (2012) it is currently not plausible that all games could be created to be equally playable and enjoyable for every human being. The current state of technology will not permit it. However, by developing new design principles and utilizing new technology it may be possible to take some steps towards that goal. They further state: "In short, we need to work to get every title to have the broadest audience possible and make sure that, for those left out of a particular title, there are other titles waiting for them to play" (p. 9).

There are a few different types of games visually impaired gamers tend to play (The Accessibility Foundation, 2015):

- Text-based games - made accessible to blind players via screen readers.
- Mainstream video games - some visually impaired gamers use audio cues in mainstream video games to play, e.g. fighting games where different punches, kicks etc. can be identified via their sounds.
- Audio games - games with only auditory and no visual output.
- Video games made accessible through modification - games that are modified by gamers themselves to be accessible (also known as "mods").
- Video games that are accessible through original design – the category that best describes the game presented in this paper.

Text-based games and mods can be fully playable and even be designed specifically for visually impaired players. However, these types of games generally lack the production values of mainstream games (The Accessibility Foundation, 2015). As for audio games and games that are accessible to visually impaired gamers through original design, there are a few examples, for instance Bit Generations Sound Voyager (Skip Ltd., 2006) and Terraformers (Westin, 2004). The most acclaimed audio games on the market are probably the Papa Sangre games (Somethin' Else 2011; 2013). In Papa Sangre, graphics is not used to convey any information and the game is fully accessible to visually impaired players. Several previous studies (e.g. Ekman, 2007; Oren, 2007) have reported that many sighted players experience difficulties navigating an audio-only interface, and audio-only games may therefore not attract a sighted audience.

2.2 Immersion

The study presented in this paper focuses on player immersion. A decade ago, Brown and Cairns (2004) pointed out that the term immersion is used inconsistently in the literature, and unfortunately this is still true. There are related concepts such as presence (Witmer & Singer, 1998), flow (Csikszentmihalyi, 1990), cognitive absorption (Agarwal & Karahana, 2000) and user experience (Brockmyer et al., 2009) that all have contributed to the understanding of immersion. Immersion can in many ways be seen as a general concept that incorporates many of the more specific terms. Weibel & Wissmath (2011), for example, states that presence relates to spatial immersion - the sense of being in the game. Flow, on the other hand, relates to immersion in the task - the sense of being absorbed by the game task. Jennet et al. (2008) acknowledge the similarities between immersion and flow - both relate to how a person can become absorbed by an activity - but also highlight the difference that flow is an extreme experience while immersion is more of a continuum. In this way flow can be seen as the extreme end of immersion. Jennet et al. propose a definition of immersion to be specifically used for studies of computer gaming experience: "immersion is concerned with the specific, psychological experience of engaging with a computer game" (p. 643). This is also how immersion should be understood in this paper.

2.3 Instruments to Measure Immersion

Numerous questionnaires have been developed that can be used to evaluate the immersion of a gaming experience. To give a few examples, Brockmyer et al. (2009) propose the *Game Engagement Questionnaire* (GEQ) based on previous work on absorption, flow, presence and immersion. Qin et al. (2009) present a questionnaire focused on studying the immersion in the narrative of a game. Calvillo-Gámez et al. (2010)

present the *Core Elements of the Gaming Experience Questionnaire* (CEGEQ), which is based on a grounded theory approach. No single scale has been established as a norm (Nordin et al., 2014), which makes comparison between studies problematic. The diversity of questionnaires is many times motivated by the context in which they have been developed. As an example, the GEQ was developed to analyse the experience of playing violent video games and has therefore a question on how scared subjects felt.

One of the most used instruments is the *Immersive Experience Questionnaire* (IEQ) proposed by Jennett et al. (2008). IEQ exists in several versions with variations in questions and scales. The version used in this paper is the modified version (Jennett, 2010, Appendix 7) which consists of 31 questions that should be answered on a 7-graded scale. The IEQ Immersion score can be computed as the sum of the 31 questions (0-6) where negatively stated questions (i.e. where high numbers indicate low immersion), have been inverted. The questions have also been clustered according to five immersion factors: cognitive involvement; real world dissociation; emotional involvement; challenge and control. A score can be computed for each immersion factor according to the same principle as the total immersion score.

The problem with IEQ and some other similar instruments, for example CEQEQ (Calvillo-Gámez et al., 2010), is that they are based on assumptions that the game contains graphics and/or that the subjects are able to use visual information. Hence, these instruments exclude visually impaired subjects. As an example, in the CEQEQ 6 out of 38 questions require subjects to be able to see in order to answer. The score can be computed by excluding questions but it may affect the validity of a comparison if one group has 20% more questions than the other. In the IEQ (Jennett et al., 2008) there is only one question, out of 31, that assumes that subjects are able to see. This question is to what extend subjects appreciate the graphics and imagery. This is relevant to most games - given that there are graphics and the subject is able to see it. Interestingly, the IEQ does not have a corresponding question related to the audio. This is a clear weakness of the IEQ – the importance of audio to the immersion of games is previously documented (Huiberts, 2010). An even bigger weakness with the IEQ related to inclusiveness is that the immersion score is computed by adding the response to individual questions to a sum. In this way the total immersion of the visually impaired is instrumentally lower when the question regarding graphics is excluded. In other words, visually impaired players are defaulted to have the lowest possible appreciation of the graphics and imagery. A better approach is to compute the immersion score using some statistical metric, such as median, that makes comparison possible and fair even when the number of questions varies.

The conclusion from this is that it is not possible to use IEQ unchanged in the context of this paper. When the experience of sighted and visually impaired players should be compared the instrument itself has to be inclusive.

3. THE GAME

The game developed in this project is a point-and-click adventure game for iOS and Android devices, designed to include blind and visually impaired gamers without alienating the sighted audience. A lot of effort was put into making it as close to a commercial game as possible with regards to production values. To include blind players in a classic graphical adventure game, design tools such as auditory icons and voice-over were used. Auditory icons can be described as “everyday sounds that convey information about events by analogy to everyday sound-producing events” (Encelle et al., 2011, p. 125),

Following the tradition of other games in the genre, such as The Secret of Monkey Island (LucasFilm Games, 1990), the game is an interactive story with a protagonist that corresponds to the playable character and gameplay mainly consisting of puzzle challenges, interactive dialogues and exploration (Adams, 2014).

The dialogue system follows the traditional branching tree concept, with a list of dialogue options displayed on the screen from top to bottom. Similar to other games in the genre, the dialogue takes place in a separate gameplay mode (Adams, 2014) in which the camera zooms in on the non-playable character as if he or she was standing face-to-face with the playable character (see Figure 1, right). To facilitate voice-over (read by the lead actress), the player can only choose among a range of topics rather than complete utterances. This adjustment also ensures that sighted and visually impaired players are provided with equal amount of information. The voice-over is activated when the player holds a finger over one of the options, but there is some delay before the voice over starts so players that have access to the graphics might never discover this feature in the game.

Figure 1. Screenshots from the game illustrating the two main gameplay modes. A user can search for, and activate interactable objects in the exploration mode (left) and select dialog options in the dialog mode (right)

By convention there are also items in the environment that the player can pick up and sometimes use for solving puzzles. Typically, this would require an inventory, a well-established game feature for storing collectible items. However, to implement an inventory that both groups of players could manipulate equally intuitively turned out to be very difficult to achieve. Instead, the player can pick up an item in the game environment that will be used automatically when the circumstances are right.

Interactable objects in the game world are revealed both graphically and audibly. Graphically by a circle surrounding the object when the finger is pointing at it, and audibly by a unique diegetic (Huiberts, 2010) auditory icon that can be traced by dragging one finger across the screen and listening in the 3D space. An interactable door, for example, sounds like someone's jerking the handle. In addition to these, the game includes extra audio feedback for every action and interaction possible.

4. THE STUDY

The game has been evaluated through formal user tests where subjects have been asked to play the first chapter of the game and then answering a questionnaire. In total, 29 subjects participated in the test, divided into three groups: sighted subjects playing with graphics (SG) and sighted subjects playing without graphics (SNG) both consisting of five men and five women each, and visually impaired subjects (VIG) consisting of five men and four women. The ages of subjects vary between 15 to 39 years of age (average age is 22.5). There is also a variation in subjects' previous experience of using games and previous experience of playing and recording music. Of the visually impaired subjects the degree of impairment varied: some were blind from birth, some could distinguish light from dark and some could read the text on the screen.

The gaming device used during the tests was an iPod touch 5G, with a 4" touch display. The headphones were a pair of AKG 240 Studio. All responses from subjects were transcribed directly by the experiment leader. No audio or video recordings were made. As a compensation for participation all subjects were given a €15 iTunes gift card.

The players had a maximum of 25 minutes to complete the chapter. During the play session, the progress of the subject was monitored and if a subject spent three consecutive minutes without making any progress, the experiment leader provided some hints. The nature of the hints given depended on the type of obstacles a subject experienced. It could for example be related to the game logic (e.g. to search for items), the feedback system (e.g. to listen for interface sounds) or the subject actions (e.g. to make slower swipes on the touch screen).

The progress subjects made in the game has been evaluated in terms of how far they reached in the *Critical Path*. The critical path consists of 31 steps of which the first four are tutorial steps before the actual game begins. These 31 steps are the absolute minimum number of interactions that has to be completed to finish the chapter.

Statistical analyses have been performed using SPSS version 22.

4.1 The Instrument Used

We have developed an adjusted version IEQ, which handles the weaknesses discussed earlier. It has also been adjusted to handle some of the problems discovered in a pilot testing. These problems relate to how the questionnaire is used and presented to subjects in a study including both visually impaired and sighted subjects. A pen and paper version of the questionnaire can obviously not be used. A computer based version is an alternative, but requires a deep understanding of what type of tools all involved subjects use. Instead, we decided to use an oral version where the experiment leader read the questions and subjects respond orally. In this way the application of the instrument is identical to all groups. The pilot test however revealed that it was time consuming to read questions and labels for the endpoints of the scale. In addition, subjects expressed hesitation on the order of the scale and the labelling of endpoints. To avoid misunderstandings and to reduce the time to present the questions, we decided to have the same labelling for all questions. All questions starts with “to what extent did/were you...” and the subjects are instructed to answer with a number between 1 and 7 where 1 denotes “to a very low extent” and 7 denotes “to a very high extent”.

The final instrument contains 18 questions, shown in Table 1, of which 14 are directly taken from the IEQ and 2 are adjustments for the specific type of game (Q12 & Q13). IEQ was created to have pairs of questions with the same content phrased twice - one time with a negative wording and one time with a positive wording (Jennett et al., 2008). The reason for this was to control the wording effect. For example, both "To what extent did you find the game challenging?" and "To what extent did you find the game easy?" are included in IEQ. In the adjusted version, used in this paper, we have reduced the number of questions by removing this kind of redundancy.

Two new questions have been added (Q15 & Q16) and they relate to how subjects appreciate the audio and the narrative. The question regarding graphics and imagery (Q14) is included in the questionnaire, but is only posed if a subject has experienced the graphics. The responses to Q4 and Q6 are negated when the immersion score is computed.

Table 1. The questions used in the immersion questionnaire. Q14 is posed if a subject has experienced graphics.

Immersion question
Q1: To what extent did you feel you were focused on the game?
Q2: To what extent did you put an effort into playing the game?
Q3: To what extent did you feel you were trying your best?
Q4: To what extent did you notice events taking place around you?
Q5: To what extent did you find the game challenging?
Q6: To what extent did you feel that you just wanted to give up?
Q7: To what extent did you feel motivated while playing?
Q8: To what extent did you feel like you were making progress towards the end of the game?
Q9: To what extent did you feel emotionally attached to the game?
Q10: To what extent were you interested in seeing how the game’s events would progress?
Q11: To what extent did you feel as though you were moving through the game according to your own will?
Q12: To what extent did you feel you were able to locate objects that were interactable?
Q13: To what extent did you feel you were able to re-locate objects that you had previously discovered?
Q14: To what extent did you enjoy the graphics and the imagery?
Q15: To what extent did you enjoy the audio and the music?
Q16: To what extent did you enjoy the narrative?
Q17: To what extent did you enjoy the game as a whole?
Q18: To what extent would you like to play the game again?

As discussed above, the immersion score in IEQ is computed as a sum. This makes the maximal score higher for sighted players as the question regarding graphic and imagery is excluded for players who have not seen the graphics. In this study, the immersion score is instead computed as the median of the answers. An alternative would be to compute the average, which is used in many other questionnaires (Weibel & Wissmath, 2011; Fu et al., 2009). The treatment of Likert scale data is however controversial. The debate (see e.g. Norman, 2010) relates to what type of statistical operations that may be applied to data that is of ordinal type. In this paper we have chosen to take a conservative approach and use statistical methods that are accepted and well established to be used with ordinal data.

4.2 Result

All 29 subjects completed the test. Of these, 16 finished the whole chapter. There were significant differences between the groups in terms of the progress they made in the game. Only 33% from the SNG finished the chapter, compared to 44% of the VIG subjects and 90% of the SG subjects. The median critical paths for the groups were 22 (SNG), 25 (VIG) and 31 (SG).

We see two potential effects that a difference in performance can have on the experienced immersion:

1. The audio-based interaction makes the task more frustrating to the player, in which case the immersion of the game could be negatively affected.
2. The audio-based interaction makes the task different but not frustrating, in which case the immersion could be affected in any direction

This section presents the results of the immersion questionnaire in order to conclude what effects the differences in progress had on the perceived immersion.

4.2.1 Immersion Score and Immersion Factors

The immersion score was computed as the median of the 18 questions (17 for players who did not see the graphics). Figure 2 shows the boxplots of the median immersions for the three groups.

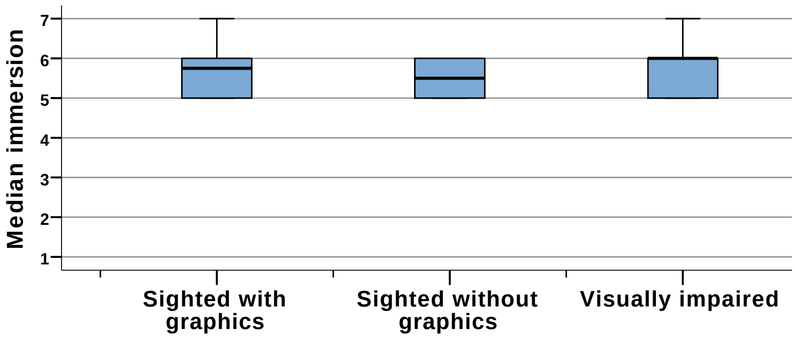


Figure 2. Boxplots of the immersion score of subjects in the groups: sighted with graphics (n=10), sighted without graphics (n=10) and visually impaired (n=9).

As can be seen in Figure 2, there are only small differences in immersion between the groups. The median score varies between 5.5 (for SNG), 5.75 (SG) to 6.0 (for VIG). The differences are not significant (p=.712 using a Kruskal-Wallis H test). Note that the lower and upper quartiles are the same for all groups. It is clear that the difference in progress has not had any obvious effect on the total immersion of subjects. Moreover, the immersion is quite high for all groups, which indicates that the game has been successful in providing an immersive experience to players. No subject had a median immersion below 5.

As suggested by Jennet et al. (2008), the immersion questions could be grouped according to five factors. Table 2 shows the immersion factors for all groups together with the result of a Kruskal-Wallis H test for each factor.

Table 2. Comparison of the immersion factors between the tree groups using a Kruskal-Wallis H test.

Immersion factor	Sighted with graphics	Sighted without graphics	Visually impaired	$\chi^2(2)$	p
Cognitive involvement (questions 1, 2, 3, 5, 7, 8, 17)	5.5	5.5	5.0	.367	.832
Real world dissociation (question 4)	5.0	6.0	6.0	1.451	.484
Emotional involvement (questions 7, 9, 10, 17, 18)	5.5	5.0	6.0	1.379	.502
Challenge (question 5, 6)	5.0	5.5	5.5	1.325	.516
Control (question 8, 11, 12, 13, 14, 15, 16)	6.0	4.5	5.5	13.636	.001

The first four factors have small variations between the groups and there are no differences that are significant. There is a tendency that SNG and VIG have higher real world dissociation. This seems reasonable as SG may be more aware of real world events through visual observations during gameplay.

For the last factor, *control*, there is a significant difference between the groups. This difference is in line with the progress of the groups where SNG made less progress than both VIG and SG. It is important to note, however, that the difference is isolated to the control factor and that there seems to be no negative effect on other factors. For example, on emotional involvement, SG has a median value which is lower than that of VIG.

This result shows that differences in progress between the groups have made the immersion slightly different. The hypothesized effect that the slower progress would cause frustration that ruins the immersion cannot be observed.

4.2.2 Response to Individual Questions

To explore potential differences between groups on individual questions, a Kruskal-Wallis H test was conducted. This revealed a significant difference between groups on question 8, 12, and 13. The biggest difference is on question 13, which relates to the ease of relocating objects that has been previously discovered. The Kruskal-Wallis H test showed that $\chi^2(2) = 17.951$, $p < .001$, with a mean rank score of 22.50 for sighted players with graphics, 7.00 for sighted players without graphics and 15.56 for visually impaired players. The other two questions with significant differences, question 8 (p=.002) and question 12 (p=.020), relate to the progress and the ease of locating new objects in the game. The boxplots of the three questions for the three groups are shown in Figure 3.

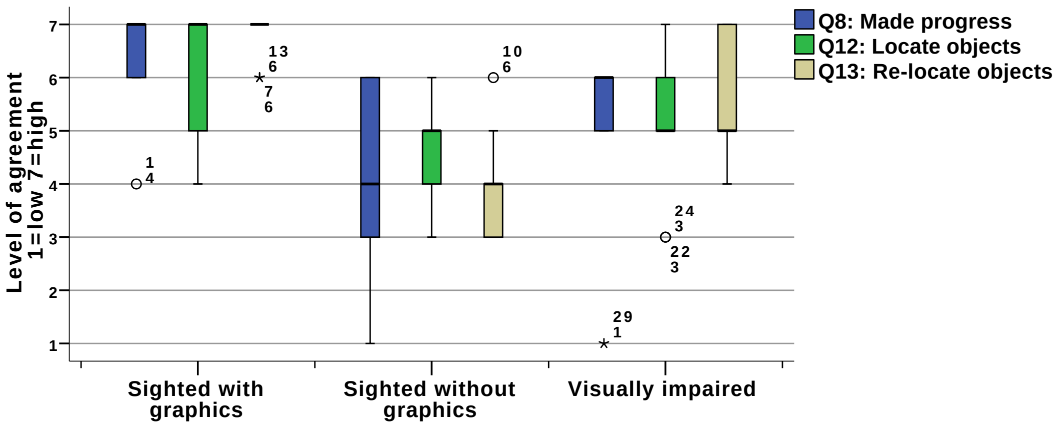


Figure 3. Boxplots of the three questions where there were significant differences between the groups.

Note that all three questions are included in the *control* factor discussed above for which there were significant differences between the groups. Figure 3 confirms the observation that it is mainly subjects from the group SNG that experience problems with locating objects and progressing in the game. This is in line with observations made in several previous studies (Ekman, 2007; Oren, 2007). Sighted players are in general not used to rely solely on listening and there is a substantial barrier to cross for them to be able to enjoy audio-based games.

Although this game is not intended to be played as an audio-only game by sighted players, it appears that they still appreciated the experience. The cognitive and emotional involvement does not differ from the other groups. This is also clear when comparing the median scores for the responses to question 17, which concerns how much the subjects appreciated the game as a whole: 5.5 (for SG), 5.0 (for SNG), and 6.0 (for VIG).

4.2.3 Other Factors that May Affect Immersion

The ability to experience the graphics of the game is not the only factor that may affect immersion. Other factors, such as age, gender, previous experience of gaming and audio, may impact on the result. The result of an analysis of how these factors affects immersion, however, revealed no significant differences between groups for any factor.

5. CONCLUSION

This paper presents a study where the same mobile game has been played by subjects with very different prerequisites – ranging from experienced gamers playing it as a regular game with graphics, to subjects with very limited previous gaming experience who played it blindfolded with audio as the only feedback mechanism. The aim with the game is to enable visually impaired and sighted players to have a shared experience. In the study, three different groups were compared with respect to the progress they made in the game and their level of immersion when playing the game. The three groups were sighted playing with graphics, sighted playing without graphics and visually impaired. All groups were mixed with respect to gender, age and previous experience of games and audio. The instrument used to observe immersion was a version of the Immersive Experience Questionnaire (Jennett et al., 2008) adjusted to be used for audio-based games and visually impaired subjects. The result shows that although there was a difference in progress between the groups, it had limited effect on the immersion. The only factor that showed a significant difference between the groups was *control*, i.e. how subjects perceived that they could control and interact with the game. The interaction with the game, when it is played using graphics, is very close to the conventions of its genre – point-and-click adventure games. The interaction with the game, when it is played using only audio, is novel and utilises diegetic auditory icons where left to right panning and volume is used to help players locate objects. Naturally, it takes longer time to locate interactable objects on the screen using audio only, but the result from this study indicates that it does not generate frustration that affects the overall immersion. This is an important result as it shows that it is possible to create games that are inclusive - sighted and visually impaired players are able to, and enjoy playing the very same game. And importantly, the visually impaired are able to play without the aid of external tools or complex mechanisms that may interfere with the gaming experience.

It would be interesting to explore the potential of extending the audio-interaction by introducing binaural sound feedback (Rumsey & McCormick, 2009), which may provide new opportunities and may improve the precision in the navigation.

No significant differences with respect to background parameters, such as age, could be found in the presented study with regards to immersion. However, there are some indications that there are differences in performance depending on previous gaming experience and gender, but the number of participants in each main group is too small to be able to make a meaningful comparison of subgroups. We are currently exploring some of these aspects in a study that focuses on an isolated part of the audio-based navigation used in the game presented in this paper.

ACKNOWLEDGEMENT

This project has been co-funded by The Swedish Post and Telecom Authority and the University of Skövde.

REFERENCES

The Accessibility Foundation, 2015, *Gaming with a Visual Disability*. Available online: http://www.game-accessibility.com/index.php?pagefile=visual [last accessed May 8, 2015]

Adams E., 2014. *Fundamentals of Game Design*. Pearson Education, New York, NY, USA.

Agarwal, R. and Karahanna, E., 2000, Time Flies when You're Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage. *MIS quarterly*, Vol 24, No. 4, pp 665-694.

Barlet, M.C. and Spohn, S.D., 2012. *Includification*. The Able Gamers Foundation, ISBN-13: 978-1479289356.

Brockmyer, J.H., et al., 2009. The Development of the Game Engagement Questionnaire: A Measure of Engagement in Video Game-playing." *Journal of Experimental Social Psychology*, Vol. 45, Nr. 4, pp 624-634.

Brown, E. and Cairns, P., 2004. A Grounded Investigation of Game Immersion. *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, USA, pp 1297-1300.

Calvillo-Gámez, E.H. et al., 2010. Assessing the Core Elements of the Gaming Experience. In R. Bernhaupt, ed., *Evaluating User Experience in Games*. Springer, London, UK, pp 47-71.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper Perennial. London, UK.

Ekman, I., 2007. Sound-based Gaming for Sighted Audiences – Experiences from a Mobile Multiplayer Location Aware Game. *Proceedings of the 2nd Audio Mostly Conference*, pp 148–153.

Encelle, B. et al., 2011. Annotation-based Video Enrichment for Blind People: A Pilot Study on the Use of Earcons and Speech Synthesis. *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*. ACM, New York, NY, USA, pp 123-130.

Fu, F-L. et al., 2009. EGameFlow: A Scale to Measure Learners' Enjoyment of E-learning Games." *Computers & Education, Vol.* 52, No. 1, pp 101-112.

Huiberts, S., 2010. *Captivating Sound: The Role of Audio for Immersion in Computer Games.* PhD Dissertation, Utrecht School of the Arts, Netherlands.

Jennett, C. et al., 2008. Measuring and Defining the Experience of Immersion in Games. *International Journal of Human-Computer Studies*, Vol. 66, No. 9, pp 641-661.

Jennett, C., 2010. *Is Game Immersion Just Another form of Selective Attention? An Empirical Investigation of Real World Dissociation in Computer Game Immersion.* Doctoral Thesis, University College London.

LucasFilm Games, 1990. The *Secret of Monkey Island*. Computer game.

Nordin, A. et al., 2014. Too Many Questionnaires: Measuring Player Experience Whilst Playing Digital Games. *Seventh York Doctoral Symposium on Computer Science & Electronics.*

Norman, G., 2010. Likert Scales, Levels of Measurement and the "Laws" of Statistics. *Advances in Health Sciences Education, Vol.* 15, No. 5, pp 625-632.

Oren, M.A., 2007. Speed Sonic Across the Span: Building a Platform Audio Game. *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, ACM, New York, NY, USA, pp 2231-2236.

Qin, H. et al., 2009. Measuring Player Immersion in the Computer Game Narrative. *Intl. Journal of Human–Computer Interaction, Vol* 25, No 2, pp 107-133.

Rumsey, F. and McCormick, T., 2009. *Sound and Recording: an Introduction*. Focal Press, Oxford, UK.

Skip Ltd., 2006. *bit Generations Soundvoyager.* Game Boy Advance game.

Somethin' Else, 2011. *Papa Sangre*. Mobile game.

Somethin' Else, 2013. *Papa Sangre II*. Mobile game.

Weibel, D. and Wissmath, B., 2011. Immersion in Computer Games: The Role of Spatial Presence and Flow. *International Journal of Computer Games Technology,* Vol. 2011, pp 1-14.

Westin, T., 2004. Game Accessibility Case Study: Terraformers – a Real-Time 3D Graphic Game. *Proceedings of the 5th International Conference on Disability, Virtual Reality and Associated Technologies, Oxford, UK, pp 95-100.*

Witmer, B.G. and Singer M.J., 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire." *Presence: Teleoperators and Virtual Environments, Vol.* 7, No. 3, pp 225-240.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

BETWEEN THE VIRTUAL AND THE PHYSICAL: MENTAL –EMOTIONAL EXPERIENCES OF SPACE IN DIGITAL GAMES AND VIRTUAL WORLDS

Samuel Gallastegui
University of the Basque Country

ABSTRACT

In the computer-based experience of space in digital games, along with the physical and the virtual, a third space is emerging, which this paper suggests to be a hybrid mental-emotional space. This paper will focus on this hybrid space from the point of view of art. This research addresses the question of how a change at interface level transforms the mental-emotional experience of space in video games and virtual worlds. The thesis is that from specific changes made at this level, different mental-emotional experiences of space in video games could in fact be generated. The aim is to analyze how these changes that are being made at interface level in the art field impact the spatial experience. In order to achieve this aim, firstly, a conceptual framework will be developed regarding the different ways interaction emerges in the mental-emotional space; secondly, an analytical model for space involvement will be proposed. Lastly, a case study will be analyzed and some conclusions offered. This paper will contribute to the definition of the features of this mental-emotional space and to the analysis of the new possibilities for experiencing space in digital games and virtual worlds.

KEYWORDS

Digital Games, Virtual World, hybridization, immersion, emersion, involvement.

1. INTRODUCTION

Miguel de Cervantes in his “Don Quixote de la Mancha” explores just how powerful an interface the printed book could be, a medium that was already a century old when he sat down to write his particular magnum opus. Don Quixote is a perfect example of a text in which the space of fiction overlaps with the space of the reader. In the first part of the aforementioned novel, the strategy of the interface (text inside text) is to immerse the reader in the space of fiction, whereas in the second part (text out of text) Don Quixote’s actions are dragged into the space of the reader. This issue is illustrated in chapter 62, in which Don Quixote arrives at a printing press workshop in Barcelona and comes across Avellaneda’s ‘false’ novel about himself.

While Cervantes in his first publication was wondering what would happen if someone actually lived the fictional life of knights and knighthood described in books in real life, in the second part he seems to be wondering what would happen if the rest of the people saw Don Quixote as a real member of society, as if fiction and reality coexisted.

I would like to suggest the idea that in Don Quixote the space of Spanish reality of XVI century could be working as an extension of the fictional world, whereas the fictional world was an extension of everyday reality in the days of Don Quixote.

Nowadays, people are able to see giants instead of mills, perhaps because they look at them through an augmented reality device. May we ask ourselves: do they get involved in the space of the giants or do they stay in the scope of the mills?

Let us try to give an answer. This paper will build a conceptual framework in order to understand the main aspects of the spatial experience of video games. Afterwards the main question, the objectives and the thesis will be formulated. Finally, to obtain a more profound exploration, an analytical model will be proposed with which I will analyze a case study and suggest some changes at the interface level.

ISBN: 978-989-8533-38-8 © 2015

2. CONCEPTUAL FRAMEWORK

2.1 Between the Physical and the Virtual: The Subject

2.1.1 Physical and Virtual Space Hybridization

Nowadays there are diverse examples of hybridization between the physical world and the space of virtual worlds, such as Simulation Based Learning (SBL), Augmented Reality (AR), or Mixed Reality (MR), among others. The boundaries between virtual space -or synthetic worlds- and physical space are blurring and the dichotomous approach between real space and virtual space has been widely questioned (Lehdonvirta 2010). Moreover, spatial hybridization is extensively used in technologies capable of adding a layer of virtuality over the physical space, such as location based and Augmented Reality devices. While visual and structural construction in virtual worlds are influenced by the physical ones (Manovich 2001) it seems to be assumed that video games generate new scientific and educational realities by means of the gamification, and new social realities by means of the increasing implementation of serious games. In addition, Castronova (2005) describes that while some synthetic worlds are closed, others are influenced by what happens in our non-synthetic world.

From an ontological point of view, Pierre Levy (1999) creates an analytical framework and states that virtualization is a process which generates a common world for individuals and includes every mediation that allow us to communicate. The Actual, the Real, the Virtual, and the Possible are four instances in the coming to pass of any phenomenon. It may be argued that it is difficult to identify clear boundaries between the Actual and the Virtual externally to the subject's experience. Therefore, this paper will put emphasis on how space is experienced rather than on the blurring boundaries between physical and virtual space.

2.1.2 Mental-Emotional Space

As an experienced gamer (and virtual world user) I may use a simple adverb of location such as "here" or "there" to locate the experience of my action. The first one associates with my physical space and the second one with the space of the virtual world. This answer goes beyond the common, but not accurate dichotomy between virtual and real, because, as stated before, both spaces are real (Lévy 1999). Again, if boundaries between physical and virtual are blurring, this is because there are limits, and therefore a separation still exists. Nevertheless, what determines my way of experiencing game space has nothing to do with either of these two spaces, but with an experiential circle of motivation, cognition and emotion (Takatalo 2011). This research suggests that this experience constitutes a third space between Physical and the Virtual Space: Mental-Emotional Space.

- Physical space, where the subject's body resides, related to the material and actual. It is structured by physical boundaries and its perception depends on the subject's senses. Its constitution is tied to the relationship of the body with the space by means of the actions that the body does in it.

- Virtual environment/space, where the projection of subject's body (avatar) moves and is related to the immaterial and the virtual. It is made out of data and its representation and interaction depends on the interface. Its constitution is tied to the relationship of interactivity with spatial representation. Interactivity activates the space, but it is straightly dependent on the subject.

- Mental-emotional space, where the subject places his experience. Experience which has many aspects: perceptive, cognitive, aesthetic, social, cultural, emotional… and are inseparable from each other. This space is structured by subjective symbolic systems that relate to the above mentioned aspects. This spatial idea is close to the concept of "magic circle" described by Montola and Stenros (2009) as a contractual barrier that separates the events and motivations of the game from those that belong to the realm of ordinary life.

2.1.3 Spatial Involvement

Digital games are psychotechnologies, term coined by De Kerckhove (1997) to address any technology that emulates, extends or amplifies the power of our minds, modifying our consciousness. He affirms that these technologies will turn our exterior world into an extensive consciousness, taking us from being "Homo Theoreticus" to be "Homo Participans". Similarly, the experience of digital games and virtual worlds is determined by the experience of taking active part in a virtual environment. These ask players to adapt themselves to -and form a relationship with- physical and social aspects of digital games, while involvement, presence and flow characterize the experience of acting in a virtual environment.

Calleja (2007) exposes a multidimensional analytical model for involvement that constitutes six frames structured on macro and micro-involvement. The macro explores motivational attractors during a long period of gaming, while the micro explores the depths of the moment by moment experience. Calleja asseverates that the subsequent intensification of experience is represented by 'incorporation', as a 'state of deep involvement that results in a shortening, or disappearance of distance between player and game environment'.

In this paper involvement is understood as the subject's adaptation to the spatial symbolic system of a particular video game, accepting a group of interactional and representational conditions that acquire consistency and meaning of space. Spatial involvement depends on how the subject-player perceives, feels and understands differentiated spatial symbolic systems, on how he or she adapts to them and keeps them apart in his or her mind. For example, a player can be completely involved in the symbolic system of a game's world and meanwhile carry out actions such as drinking, eating or speaking. These actions belong to the logic of physical space, and have different relationship between motivation, cognition and emotion.

2.2 Towards the Physical and the Virtual: The Interface

2.2.1 Transferring Spatial Involvement: Narrative, Representation and Interaction

This paper proposes that in digital games and virtual worlds, involvement is transferred by representation, interaction and narration. Just as when walking down a crowded street the subject receives and deals with kinaesthetic information by using sidewalks, crosswalks or subways, when navigating throughout the space in a virtual world the subject uses buttons, commands, toolbars to deal with that kinaesthetic information that configures the space (perspective, sounds, maps, textures, colors...). This is because representation and interaction are two sides of the same coin that it is used to transfer the user's involvement.

On one side the transfer is given using images and gestures, and on the other side using rules and instructions. Narration is the coin itself, or, better said, how it is ordered and configured.

2.2.2 Orientation of the Interface

De Souza e Silva states that the universalization of ubiquitous technologies has enabled the existence of hybrid spaces where "users do not perceive physical and digital spaces as separate entities, and do not have the feeling of 'entering' the Internet, or being immersed in digital spaces" (de Souza e Silva 2006). The subjects situate themselves in space according to where their actions are involved and make sense. Let us study an example. The game "Dance Star Party" takes place in a very simple Virtual world where the player follows instructions and performs a dance which is required by the game. The dance is a reality that happens in the player's living room but it is involved in and makes sense within the game, therefore we could say that the living room has turned to be part of the virtual world and that the involvement is orientated towards the physical world. Depending on where the interface orients the involvement of subject actions a taxonomy can be made:

- Immersive orientation. The mental-emotional space of the subject is engrossed by a synthetic virtual space, his awareness of physical self is immersed inside a different environmental space and all the subject's actions are oriented towards that virtual space. This kind of orientation is implemented in most main stream games.

- Emersive orientation. The subject's mental-emotional space is pulled to the physical space, where he or she bears awareness of the physical self. All the subject's actions are oriented towards the physical space but mediated though a virtual environment. This sort appears especially in games that use such controllers as Wiimote or Kinect, or those that use location aware device to use physical space as a playground.

- Comprehensive orientation. The subject's mental-emotional space is expanded by a synthetic virtual space but also takes place in the physical space. All subject's actions are oriented both towards the Virtual and the physical space. These are, for example, location based and Mixed Reality games.

ISBN: 978-989-8533-38-8 © 2015

Immersion can alter the way the world that is perceived by the player (Weger & Loughnan 2014).Conversely, an emersive gameplay actually alters the reality, as do Alternate Reality games, that change reality in many different ways (Haring 2010).

3. QUESTIONS, THESIS AND OBJECTIVES

The main question this paper will address is how this third mental-emotional space between the physical and the virtual space is experienced and how new experiences of space in digital games are appearing. The question is: could a change at the interface level transform the mental-emotional experience of space in digital games and virtual worlds into something different?

This study holds the thesis that making changes at the interface level would generate new mental-emotional experiences of space in digital games and virtual worlds. This paper will explore these changes and analyze how they impact the experience of space. The objectives of this research are:

- To analyze how the mental-emotional third space between the physical and the virtual is experienced, in relation to the concepts of spatial involvement and orientation of the interface.

- To explore how new experiences of space in games can be created by making changes at the interface.

4. ANALYTICAL MODEL FOR SPATIAL INVOLVEMENT

The mental-emotional third space that this paper introduces is determined by the experience of acting in a particular space, which is at the same time determined by the Spatial Involvement of those performed actions.

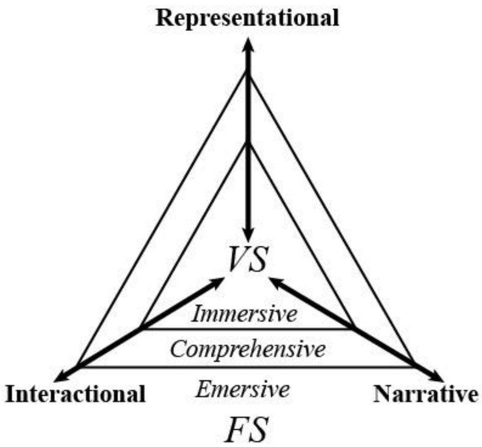


Figure1. Analytical model of interface orientation

I propose a three-dimensional analytical model for Spatial Involvement, as shown in figure 1, in which narrative dimension, representational dimension, and interactional dimension are three different areas of variation between physical and virtual space. These areas of variation indicate subject's sense of acting in a particular symbolic system, in which Narrative correlates with the happening, interactional correlates with the performing, and representational correlates with the appearance.

The maximum grade of immersion is a total independence defined by the loss of awareness of physical space (FS). The maximum grade of emersion is a complete directness outside the virtual space (VS). This model has been partially inspired by Milgran-Takemura's (1994) Reality-Virtuality continuum.

4.1 Narrative Dimension

The concept of narrative that is held in this research suggests not only a sequence of events, but also spaces, concepts, ideas or any other abstract data. Therefore, the narrative dimension designates, along with rhetorical mode of discourse, the whole arrangement of an interface. This concept of narrative is closely related to Manovich's (2001) Database Cinema. Digital games and virtual worlds reproduce the narrative tradition of visual languages mediated by a camera, but instead of registering physical phenomena, the camera in digital games gives access to a set of interconnected data and algorithms that are decoded as forms of a synthetic space.

4.1.1 Narrative Dimension Range

Narrative correlates with the happening of subject's actions, which can range from a total narrative independence of the physical world and a complete autonomy of virtual worlds, to a real-life determined narrative. E.g. U2 pop music band performing a simulated concert in Second life is determined by those that they perform in Physical Space. Here is an array of instances in the narrative dimension range, from the most immersive (virtual space orientated) to the most emersive (physical space orientated) narrative:

1. Traditional videogame. Events that occur in a virtual environment only make sense inside the fictional world, because refers only to fictional contingency. First Person Shooter games are a good example.

2. Metaverse. Although the events are virtual, the porosity of metaverses allow them to be transferred to physical life events. Events that occur in a virtual environment can become a fact in physical space, as when Anshe Chung became not only a virtual world millionaire but also a real-world one too.

3. Ubiquitous game. In these games the fact of exploring physical space gives information of virtual environment because it has been covered by virtual layers. They are usually provided with GPS devices.

4. Serious game. These have a purpose beyond mere fun that belong to real-life and physical space, such as educate, train or inform (Michael & Chen 2006). For example, America's Army has been reported to be an utterly efficient recruitment tool for the United States Army.

5. Serious ARG. Alternative Reality Games use real life events and physical space to develop transmedia networked narrations that are carried out collectively. The main philosophy is refusing the fact of being playing a game, which is called 'This is not a game'(Szulborski 2005). The serious ARG are those that have a purpose beyond mere fun. World Without Oil was a game that asked players to imagine the world as if a real oil crisis were occurring, and players had to come up with ideas to overcome the situation.

6. Gamificated process. As Deterding (2011) defines it "is the use of game design elements in non-game contexts". Although transformed, gamificated processes are still more related to physical than to the virtual.

7. Reality games. These are games that are immersed in real-life and can be mediated as digital games, such as true or dare. These games refer to real-life events and are exclusively related with physical space.

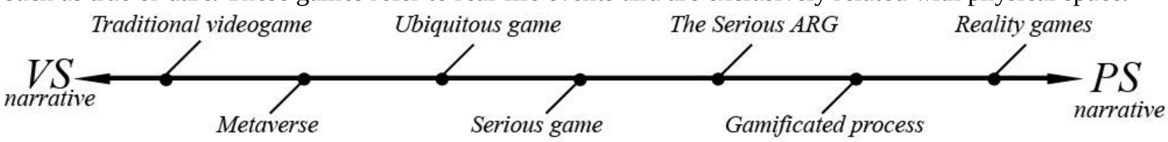


Figure 2. Cases ordered from the most immersive to the most emersive in relation to the narrative dimension.

4.2 Representational Dimension

The procedures of representation are connected with a long cultural tradition. Cartography, descriptive geometry, renaissance painting, scenic space, cinematographic space and virtual reality define the current representation of the space in digital games and virtual worlds. As Manovich says: "The visual culture of a computer age is cinematographic in its appearance, digital on the level of its material, and computational (i.e. software driven) in its logic" (Manovich, 2001: p.241).

4.2.1 Representational Dimension Range

Both physical and virtual space have a constant influence feedback from each other in its Representational Dimension. There are several possibilities of correspondence between physical and virtual space in this representational dimension that range from a totally virtual depiction to representations of physical world.

1. **Virtual representation.** Those are representations of virtual space and objects, without any relation with the physical space. Virtual worlds are a good example of this representational mode.

2. **Augmented virtuality.** It consists of including representations of real world inside the virtual space.

3. **User interface**. This refers to graphic interface that display the options, functions or applications of any operative system or program.

4. **Virtual window (trompe l'oeil).** Those are virtual frames where the space on display is generated by computer and integrates in real space like a trompe-l'oeil, as a part of architectural space, therefore expanding the physical space.

5. **Virtualized reality.** These are virtual representation of reality, such as virtual models of real buildings, or the process of capturing actor's movement for a videogame production.

6. Augmented reality. It consists of amplifying the real objects by projecting images on them or by mixing the images taken by a camera with a digital layer. These images belong to the physical space, but as they have been modified digitally they are also virtual.

7. **Representation from reality**. Those are representations of physical space and objects.

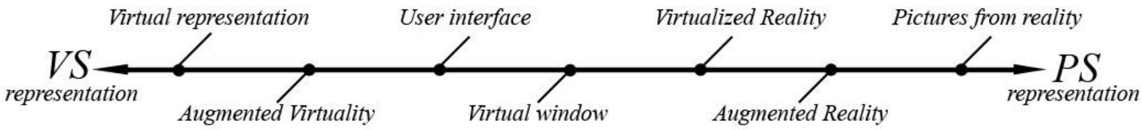


Figure 3. Cases ordered from the most immersive to the most emersive in relation to the representational dimension.

4.3 Interactional Dimension

On an operative level, the interface is a central code for the information society, a meta-tool to deal with a world of data. In addition, the interface is how the content appears, without the possibility of separation between the interface experience and its contents. Ergodic is the term coined by Aerseth (1997) to describe the narratives that need a non-trivial effort to allow the reader go through it. Digital games require ergodic actions to navigate space, such as to move a joystick, a mouse or to gesticulate in front of a computer.

4.3.1 Interactional Dimension Range

Actions performed in a virtual space are in correspondence with the physical space, and vice versa. This is a range of different devices from more virtual space oriented to the more physical space oriented:

1. **Multimodal devices** (immersive). Actions are performed directly in virtual environment and codified by head mounted stereoscopic displays with haptic systems. The player places interaction in the virtual space.

2. Simulator controller. Devices designed to give a more direct interaction with the virtual environment, where player`s actions are codified by a simulated physical interface.

3**. Human-Interaction devices**. Those designed to interact with computers, from a keypad to a gamepad.

4. Computer vision and gesture recognition. These are what Jull (2010) calls mimetic interfaces. Dynamic actions performed in physical space enable interaction with virtual space, combining both.

5. Location aware device. With a location aware device the whole planet can be turned into a playground, as happens with geocaching games.

6. Radio controller. These are used to control object in physical space, like drones, for example.

7**. Non-meditated interaction.** This is direct interaction with the object in physical space.

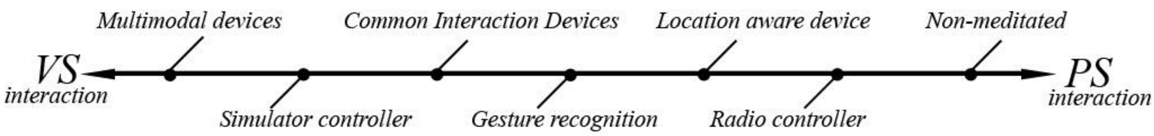


Figure 4. Cases ordered from the most immersive to the most emersive in relation to the interactional dimension.

5. A CASE-STUDY

In order to explore the questions that have been raised, this paper will apply the analytical model proposed to a case-study so as to explore how this third space emerges, which is the cemeteries and memorial parks in the virtual world Second Life (SL). There are several cemeteries in SL where residents can allocate a virtual grave to recall the memory of their dearest as well as bury their avatars that will not be operative in Second Life any more. Haverinen (2014) has studied the social online mourning as a contemporary practice that, as she states, has transform the web into a new space for mourning and honoring.

Figure 5. Picture taken in 2012 in a cemetery inb Second Life.

This picture has been taken in a cemetery in Second Life. In the foreground, on a gravestone with the form of two books the inscription “7/12/99 - 16/6/2000 an Angel on earth” can be read to recall the memory of a six months dead baby. In the background there are some other graves, one of them is for an Avatar. The symbolic implications of this virtual space can help to understand how the mental-emotional space between the physical and the virtual is experienced.

Narrative dimension

This particular virtual space refers to the memory of a dead person, which is a physical world contingency and its emotional consequences lie in the physical space symbolic system, because gravestones are funeral landmarks. Therefore, meaningful actions, such as visiting a virtual memorial space or burial space, situate the action outside the virtual world, and it cannot be understood without the physical dimension that is attached to this virtual ritual.

Representational dimension

On the one hand, the whole representation dimension in the virtual space is in great debt with a long burial representational tradition, to which elements like arched gravestones, crosses, angels, flowers, fences, walls... belong. On the other hand, the photography of the dead baby which appears in the gravestone is a mediated picture of a physical space.
As a result, it can be said that the appearance of this virtual space is a complex compilation of visual symbols which find their origin in the physical space but that actually make sense as a part of a hybrid mental-emotional experience of space.

Interactional Dimension

Navigating space, focusing the camera on an object, or constructing virtual 3D objects are codified actions performed through a keypad. The interaction model that is used in SL is closer to virtual space than to the physical. Although the subject navigates a virtual world, the action of visiting a memorial landmark, more than enclosing it to the virtual or physical worlds, places it in what this paper suggest being an emersive mental-emotional Space.

Further case-studies

There are other cases of blending life and game or physical and virtual space. *Peacemaker*, for example, is a government simulation game on Israeli-Palestinian conflict. The player has the choice to play the role of the governor of one of the two countries. As the content of the game and the pictures have been taken from the real conflict, the narrative dimension is highly emersive, although in the other two dimensions it is an immersive game. It is also worthwhile mentioning a gamification of running called *Zombies run!* that recreates a situation after a zombie catastrophe in which some runners risk their lives running among zombies to collect objects. In this case, though the narrative is immersive, the interaction with the game is emersive, for it consists of running in the streets while listening to a radio emission that informs about zombies nearby.

6. CONCLUSION

The experience of playing constitutes a mental-emotional space that is determined by many aspects (cognitive, cultural, symbolical, emotional, etc.). It can, depending on variables of narrative, representation and interaction, be placed in the virtual or in the physical space. Denying real/virtual dichotomy but admitting that there are more than two spatial realities, this paper offers the seeds of a theoretical framework to analyze how the subject's spatial perception and awareness depends on spatial involvement.

The key to new spatial realities is in the role of interface, where little changes can orient the game content emersively to the physical space or immersively to the virtual space. Therefore, new spatial modes of experience can be generated by connecting physical and virtual spaces in different ways. Nowadays, the hybridization between physical and virtual space has allowed players to get involved in the space of giants as much as they can stay in the scope of the mills. The interface prolongs the sense of the action beyond the boundaries of space, creating an emersive experience of the space. Immersion only describes partly the experience of the player. There is also an experience in which the game gets out from both its virtual and conceptual boundaries and takes place in the physical space, altering or transforming it, and transferring its virtual content into the ordinary life. That experience of being surrounded by the game instead of entering it has been proposed as emersion in this paper.

Further studies, that are part of a PhD Thesis, will focus on the concept of emersion, as a phenomenon that may be the common ground of mixed reality games, serious games or gamification. Additionally, the proposed analytical model will be developed to be applied to any particular game and determine if its emersively or immersively oriented.

REFERENCES

Calleja, G., 2007. *Digital games as designed experience: Reframing the concept of immersion*. Victoria University. Available at: http://www.gordoncalleja.com/GordonCalleja_Digital_Games_as_Designed_Experience.pdf

Castronova, E., 2005. *Synthetic Worlds. The business and Culture of Online Games*, Chicago: The University of Chigago Press.

Deterding, S. et al., 2011. Gamification: Toward a definition. *CHI 2011 Gamification Workshop …*, pp.12–15.

Haring, P., 2010. *How Alternate Reality Gaming changes reality*. VU University Amsterdam Priscilla.

Haverinen, A., 2014. Death and Mourning Rituals in Online Environments. *J@rgonia*, 24, pp.35–41.

Juul, J., 2010. *A Casual Revolution*, Cambridge, Massachusetts: The MIT Press.

Kerckhove, D. De, 1997. *The skin of culture*, London: Kogan Page.

Lehdonvirta, V., 2010. Virtual Worlds Don ' t Exist : Questioning the Dichotomous Approach in MMO Studies Abstract: Virtual Worlds Versus the Real World. *Game Studies*, 10. At: http://gamestudies.org/1001/articles/lehdonvirta.

Lévy, P., 1999. *Qué es lo virtual?*, Buenos Aires: Paidós.

Manovich, L., 2001. *The Language of New Media*, Cambridge, Massachusetts: The MIT Press.

Michael, D. & Chen, S., 2006. *Serious games: Games that educate, train, and inform*, Canada: Thomson Course

Milgram, P. et al., 1994. Augmented reality: a class of displays on the reality-virtuality continuum. *Proceedings of SPIE*, 2351, pp.282–292.

Montola, M., Stenros, J. & Waern, A., 2009. *Pervasive Games: Theory and Design*, United States: Morgan Kaufman Publishers as imprint of Elsevier.

De Souza e Silva, A., 2006. From cyber to hybrid: Mobile technologies as interfaces of hybrid spaces. *Space and Culture*, 9(3), pp.261–278.

Szulborski, D., 2005. *This Is Not A Game: A Guide to Alternate Reality Gaming*,

Takatalo, J., 2011. *Content-Oriented Experience in Entertainment Virtual Environments*. University of Helsinki.

Weger, U.W. & Loughnan, S., 2014. Virtually numbed: immersive video gaming alters real-life experience. *Psychonomic bulletin & review,* 21(October), pp.562–5.

TOWARDS LUDOACOUSTIC IMMERSION PERSPECTIVES ON TRANSCENDING THE VIRTUAL AND THE REAL OF FUNCTIONAL SOUND AND MUSIC IN INTERACTIVE MEDIA

Hans-Peter Gasselseder
Aalborg University, Department of Communication and Psychology, Musikkens Plads 1, 9000 Aalborg, Denmark

ABSTRACT

Regardless of listening to a small studio or to a huge cathedral, the environment in which sound and music is performed can act as an agent of meaning in how we relate to our actual surroundings. Applying a short reverb on a love song may point at a setting of upfront intimacy and accessibility. On the other hand, situating that song within the blurry inaccessible spaciousness of a cathedral may not improve your chances to make your date stay past dessert. In both cases expressive compatibility of room acoustics is determined on expectations generated from the music material: But does the environment as well as its functional setting in user interaction also arouse expectations towards the expressive properties of sound/music? And how might these properties affect selective attention, the sense of immersion/presence and finally the decoding of semiotic agents in the music dramaturgy of an interactive environment? In recorded sound-fx and music, expressive artifacts of space contribute to forming expectations on how an object [sound source] is interacting with an environment [room] while also accounting for the associated resonances occurring in that object. By switching attentional focus between action and resonance in object and environment as a result of comparing expectations to incoming stimuli, the context of a virtual situation is simulated by referencing to a syntax on body-object-environment interaction. This virtual syntax may be partially projected onto the situational context of the recipient leading to antecedents of immersion depending on emotional arousal and personality traits such as empathising-systemising. After having outlined a conceptual framework describing the mediation and agency detection of sonic expression within the acoustic properties of situational contexts, the last part of the paper shall provide an outlook on how these agents may be translated to meaningful structures that are yet to be studied in interactive media such as video games.

KEYWORDS

Sonic Interaction, Game Audio, Music Expression, Semiotic Ecology, Immersion, Situation Awareness, Agency.

1. INTRODUCTION

If a performance lacks of any hints of expression such as modulations of pitch, volume/dynamics, attack-release or resonances of the room, what is left to the listening experience?

Previous research suggests that varying combinations of expressive parameters in music relate to other sensory modalities as well as to more abstract concepts such as narrative and drama. These effects are assumed to take part in directing attention towards the progression within the dramaturgic arch and are often strong enough to significantly alter the interpretation of narratives (cf. Boltz et al. 2009; Bresin et al. 2011). Editing and cinematography, rhythm, legato or staccato in music accompaniment of low versus high arousing scenes form examples of how these effects can be obtained within culturally established conventions of negotiating individually valid, connotative meanings as well as herein expressed accentuations corresponding to structural and emotional features of the portrayed scene contents (cf. Gaver & Mandler 1987, Magliano et al. 1996, Vitouch 2001, Wingstedt et al. 2008). These conventions may have become a crucial part of our environment as they have been manifested not only in music-linked parameters of expression but also in everyday communication (cf. Allesch & Krakauer 2006).

Intermodal matches of expressive parameters and emotion category may be found in vocal expression as much as in any other sound source humans have encountered over evolution (cf. Juslin & Laukka 2003). Similar expressive characteristics may be found in sound effects partially due to an acquired understanding of perceptual realism (cf. Gasselseder 2012). Shorter length, high volume, volume variability, high frequency partials and shorter attack as witnessed in vocal expression of high arousal emotions or similarly in sounds created by fast passing cars exemplify that our sense of a link between sound and the cognitive-emotional system may have originated from our immediate surroundings. Following this there is reason to believe that correlates of perceptual realism may also set the ground for connotative links of music in the virtual and its representational quality of our environment in the real world.

By accepting the above-mentioned evolutionary view on sound perception, the question arises on how sound and music expression can be perceptually realistic and imply a persuasive situational context to the listener (cf. Zwaan 1999, Gasselseder 2012). The notion of fidelity, even while being traditionally linked to the representational quality of an audio source, would then extend to a recording artifact as a perceptually realistic/intrinsically valid reflection of acoustic expression in a series of events in time. As in the case of drama, the associated reception processes may depend on the experience of agency, as well as markers of perception styles and episodic memory subject to individual dispositions (cf. Eschrich et al. 2008). On account of agency detection, expressive fidelity of music or sound may give rise to the attribution on what agent is addressed in a given context and how that agent relates to the listeners' perception of the virtual and the real situation. The present paper therefor intends to describe these processes by examining the non-mediated dimensions of sonic expression in a communication chain that illustrates how an understanding of interacting situational contexts between the virtual and the real can be obtained. After having outlined a conceptual framework describing the mediation of sonic expression within the acoustic properties of situational contexts, the last part of the paper shall provide a short outlook on how these agents may be translated to meaningful structures that are yet to be studied in interactive media.

2. EXPRESSIVE FIDELITY

Returning to the introductory question on the importance of expression in music, one may suggest that a specific music instrument may not be recognisable when played without expression or contrary to the conventions listeners base their expectations on. According to that, arranging for orchestra in an unconventional manner may result in the impression of listening to artificially synthesized music while actually being presented with a live concert setting. Conversely, a conventional orchestration performed by the means of using virtual instruments allows nowadays‘ electronic composers to make audiences believe that they were listening to a recording of a real orchestra. Moreover, recent developments in functional music imply a change of convention in recording techniques and orchestration that accounts for the sonic limits inherent to virtual instruments (cf. Kallionpää & Gasselseder 2012). Such trends can also be observed in the conception of sound effects. A between-groups study by Larsson and colleagues (2005) suggests reproduction techniques [stereo vs. binaural] of room acoustic cues to play a significant role in influencing presence ratings. Nevertheless, it is noted that the implications of these findings may depend on an adapted design of stimuli representing information on localization according to linking interaction principles across modalities (cf. Ozawa et al. 2003). Residuals associated with the principles of sound design for stereo remain in more advanced reproduction techniques, such as binaural or surround renditions [e.g. in foley ignoring two-way microphone pickup angles relative to the listener]. Such transitional aesthetics may correspond to the expectations of the audience more easily as a result of artifact literacy (cf. Gasselseder 2012; Chueng & Mardsen 2002). With regards to a process of forming expectations towards authenticity it then may be hypothesized that the acquired understanding of expression depends on an interaction between the physical object and the environment rather than the acoustic fingerprint of an object per se. This interaction of 'expressive fidelity' enables the listener to give trust in the reliability of the information communicated through a filter of acoustic conventions negotiated between the recording artifact and the sound production/performance. Defining music expression in this way does not exclude the perceived authenticity of compositions that do not meet the requirement of sonic fidelity in the sense of perceptual realism, as in the case of highly popular 8-bit-scores such as 'Super Mario Bros.' (Nintendo 1983). What matters further is how well expressive fidelity matches the situation of the listener and the subject being at the focus of attention (for a review on music and situation see Juslin et al. 2008).

Accordingly, if we are listening in a small studio or in a huge cathedral, the room sound as well as the actual content of sound and music act as agents of meaning in how we relate to our surroundings [e.g. short reverb length may point at upfront intimacy and accessibility vs. long reverb lengths' blurry inaccessible quality may suggest public places]. The interaction between human expression and the physicality of the real world informs about how expression coming from an individual causal agent [such as a performer] translates to the physical world of room acoustics. Expressive fidelity, then, describes a function of the range of energy that can be invested in modifying the physical characteristics of the interaction between an object [instrument] and its environment [room] in meaningful terms of agency detection.

If we continue to hold on to the notion of two distinct interacting agents of meaning during music listening, so that individual expression of a performer [object] represents a state of mind and room response [environment] of the scenic world within a virtual space, we may ask for the accordance of this translation to the interaction between our own internal state of mind and our own surroundings in the real world. To put this in simpler terms, an analogy is drawn to a stimulus detection experiment that concerns the antecedences of distinguishing a sound from a room tone that is inherently emitted by the environment. To discern a sound from the acoustic qualities of the room may suggest an action taking place in one's surroundings. Identifying such causal agents may be of relevance to the listeners' attentional focus by generating expectations towards environmental responses to acting objects (cf. Bruner & Postman 1949; Popper & Fay 1997, Barsalou et al. 1993, Zwaan 1999). Knowing how a room will resonate to an active object allows for simulating the context of a given situation. The properties experienced in these objects may be identified by their expressive fidelity, that is the synergetic perceptual quality arising between the object-environment relationship suggested by the recording artifact and the situational context of the listener (cf. Begault et al. 2001; Larsson et al. 2005). But at what point and to what degree would subjects attribute a causal identity to the responses of the room aroused by the object? And how is attention towards an expressive object directed? Before covering these issues in some more detail it may be stated that regardless of the presence of proximal cues of expression, the attribution of the cause-effect relationship needs to be considered as an antecedence of experience as opposed to a specific form of understanding (cf. Leman 2010). Thus, it may be worthwhile to study expressive fidelity in terms of the listener's experience of situation.

3. AGENCY DETECTION AND SITUATIONAL CONTEXT

Research on non-mediation has seen several attempts to describe an experience of situational awareness within a real and virtual scenario. Terms such as 'immersion', 'presence', 'involvement', 'absorption', 'suspension of disbelief' as well as 'self-location', 'flow' and 'possible actions' form examples of overlapping constructs that sometimes are used interchangeably and merely denote a specific application, such as virtual reality or challenge-based video games. However divergently defined, it may be argued that common ground is found in the notion of specific motivational states and altered cognitive representations of situational factors. The first refers to an intrinsic as well as utalitarian source of attentional focus serving task-dependend evaluational needs. The latter concerns the environmental field of attention as to in what regard an external entity/object/place exists on its own beyond the realm of user perception. Combining both aspects of perception, users assess the reliability and credibleness of information by actively sampling sensory data and generating expectations towards hidden objects or causes that are believed to generate those data (cf. Friston et al. 2012; Gregory 1980). If this hidden data is believed to be essential to survival [e.g. by the means of reducing surprise] it will contribute to a model of a world that in turn will affect subsequent expectations and result in information filtering. In the course of on going hypotheses testing our world model will increasingly be enriched by matching information and thus materialize into perception. Similarly, one may think of our motivation to reduce surprise to translate to an increase in reliability and credibleness of the available information. Despite an object's unsurprising/credible initial appearance [as it may hold true for a 3D projection of a vase, to give an example], our brain actively searches for salient features of the world that a) are intrinsically relevant [knowing it is a projection and expecting features that describe its altered physical properties] and b) will either disclose themselves or be discovered by active sampling [e.g. over time the projection experiences flickering or the user notices the lack of tactile feedback] (cf. Friston et al. 2012). A credible representation of situation, therefore, requires an attribution of agency that is characterized by separating the content in question from the user and assigning an independent existence to it. Another requirement that follows the tradition of Gestalt psychology concerns the content's ability to affect and organize its constituting elements autonomously. Any content can be thought of to represent a system or entity of its own that goes beyond the sum of its parts (cf. Slater 2002; Brenton et al. 2005). This ability may not be limited to entities such as virtual characters but also extend to how an object's perceptual salience and temporal position is seen in relation to the environment. Representations of these so-called relational differentials that essentially negotiate the position in the field of action and ascribed attitudes of agents are assumed to play a major role in the experience of absorption and imaginary immersion (cf. Gasselseder 2012).

Making sense of behavioural interaction by attributing intentions appears to be a ubiquitous process aiming to ascribe events in the environment to agents in a particular situation. However, it has been suggested that sensory cues which cannot be assigned to a particular acting entity will still be associated with an agent that is representative of a time-domain based structuring process of meaning (cf. Gasselseder 2012). These processes or meaning structures can be linked to higher levels of episodic processing concepts such as Bruner's (1986) organizing principle of narration and drama and, in a more extreme manner, have been associated with the foundation for belief in sentient agents of superior origin such as fate or god (cf. Gray & Wegner 2010). An alternative view or terminology of these representational concepts may be considered in the notion of situational context. The unexplained nature of a twig snapping in the forest, a popular example in the literature, suggests an object being altered within an environment (cf. Blesser & Salter 2007). The cause of this event, such as wind, however, is not attributed to the environment due to its pre-defined attribute of specifying the realm of physical action in the world model. Moreover, the environment may only provide indirect cues of the action having taken place, as for example by acoustic stimuli such as reverberation, making it unlikely to be considered as the source of the event. Knowing this, our mind is prepared to assign an abstract entity to the snapping sound that is based on meaning structures that have arisen from survival strategies and allow inferring relational differentials as well as expectations towards the situation. Thus, the notion of situation implies a holistic experience of attitudes and consequences that are derived from the sub-sum of developing states of agent-environment interactions. The most important distinction of situation, though, may be seen in its inclusion of the perceiving subject per se (cf. Ruby & Decety 2001). Expectations are interpreted in a way as to detail the personal relevance of current as well as future states of agents and thus ensure adequate choice and placement of actions in order to maximize benefit in social interaction or other contexts, such as gaming performance. However, by adding non-diegetic sound and music to interactive media, additional information with regard to states of agents and personal relevance is provided to the user.

4. THE RE-EXPERIENCE OF SITUATION IN SOUND AND MUSIC

In what is heard in sound and music we acknowledge an intention initiated by an actor translated to the physical world by the means of an object/instrument and its interaction with the environment. During listening only the latter two agents are audible, leaving the actors' state of mind [or attentional focus] up to our processing of the resonating expressive vocabulary exhibited in the interaction between the object/instrument and the room. Research on embodiment as well as on synaesthesia suggests that a large portion of perception and mental activity, including music listening, can be observed in corresponding responses in the brain such as the mirror neuron system as well as in translations to physical sensations and corporeal coarticulations (cf. Janata 2009, Schubotz 2007, Zatorre & Halpern 2005). By interpreting results obtained from a series of empirical studies, Leman (2010) assumes that corporeal coarticulations allow for the feeling of agency, of being moved or activated by music. The formation of meaning, then, may be inherently linked to mental state concepts flowing into physical movements that go in accord to the cause-effect relationship as experienced in specific sound and music-linked characters of agency. Hence, in music listening the inclusion of a third distinct agent mediating the interaction between individual expression [virtual/object] and the environment [virtual/room] is suggested in the extended body, a proto-physical schematic representation that projects partials of re-experienced expressive fidelity as syntax onto the current listening situation [real/world] (cf. Jauk 2007, 2012; Leman 2010). These embodied manifestations entail the representational framework of physical properties involved in interacting with objects and environments. By feeding this framework with information on the syntax of expression as implicated by re-experiencing expressive fidelity, a link between the virtual [recording] and real [listening] situational context is established. Such a contextual link, in turn, allows for the emergence of a combined syntax of object-environment representations, which may be consulted whenever cause-effect relationships are detected to run synchronously during the contrasting of expectations to incoming stimuli.

The extent to which these projections of expressive fidelity contribute to generate expectation may depend on characteristics of the usage situation [such as sound system or multi sensory inputs] as much as individual cognitive styles. An example for the latter may be given in Baron-Cohen's (2003) ‘Empathizer-Systemizer Theory’, in which Kreutz and colleagues (2008) distinguish empathetic as well as systematic music listening styles, the first reacting emotionally stronger while the latter more attending to structural features. Systemizers may exhibit embodied cognition to a greater degree in a mimetically manner as opposed to Empathizers’ more pronounced physical sensations as a result of emotional reactions. In line with this, listeners may experience music expression differently by adjusting their expectations according to different criteria involved in the integration and elaboration processes required to keep the representational framework in sync to incoming stimuli.

5. APPLYING EXPRESSIVE FIDELITY TO SOUND AND MUSIC IN VIDEO GAMES

In view of the outlined framework experiencing expression in sonic modalities necessitates an interaction of expressive fidelity between multiple domains of situational context. This interaction spans between the situation of listening and recording in conjunction with other content of the media presentation. By projecting expressive fidelity onto the context spanning between those content [virtual] and listening [real] situations, sound and music hold the potential to alter relational differentials and absorption experience. Subsequently, these hybrids of situational context affect the building of hypotheses of perception, which are juxtaposed with incoming stimuli. In this way, expressive fidelity is assumed to indirectly influence selective attention and agency detection (cf. Gasselseder 2012, Gasselseder & Dobler 2009).

Increasing the amount of reverb on a sound clip of stepping shoes, for example, may elicit the impression of that object moving away from the listener. However, when the listener moves away from the sound source the opposite conclusion may be drawn so that the stepping shoes seem to remain on the same location in the recording despite their experienced increase of reverb amount. This suggests a simultaneous focus of attention on two distinct situation domains, the one where sound and music is played in [virtual] and the other where the listeners are situated [real]. Expressive fidelity as a quality of representing interaction syntax affords a link between the virtual and the real domain. If expectation contrasting is synchronised and met by other sensory stimuli in both domains, embodied/proto-physical representations may extend to support experiences of non-mediation by the means of providing persuasive situational context. At this, the amount of projected expressive fidelity may depend on the perceived level of congruency in relation to the remaining stimulus modalities, such as visuals [e.g. ambient light]. The most important role of music, however, may be ascribed to congruency of content expression [e.g. matching expressed arousal in music and gameplay scenario] as has been demonstrated in studies on immersive conditions of adaptive music in video games (cf. Gasselseder 2012, Gasselseder & Dobler 2009). Here, music takes influence on the updating of our world model by prompting the retrieval of schemata based on prior experiences and beliefs. Filter effects stemming from hypotheses testing contribute to allocating higher salience to incoming stimuli that match the contents of prior beliefs. This not only leads to the picking up of sensory data but also helps to uncover hidden features and promptly process semantic connotations with regards to personal relevance.

Accordingly, in recording sound and music it may be more desirable to capture expressive fidelity in high resolution [such as mixing several microphone perspectives] rather than an approximation of reality. While such approaches are commonly found in foley for film and video games, their application has traditionally been limited to providing functional realism to the audience/player. A possible extension of this concept may add a sense of internal representations attributed to varying states of the situational context. Such relational attributes towards the environment may be accomplished, for example, by altering expressive fidelity temporarily in order to mimic internal states of interacting agents. This is exemplified in the video game ‘Bioshock Infinite’ (Irrational Games 2013) where poor health of the avatar results in added reverb on external diegetic sound while mixing in internal sound cues of visceroception such as heartbeat and breath.

Whereas external diegetic sounds remain in the background, diegetic music appears to be more pronounced even when moving away from the radio that is placed in the environment. Together these expressive cues suggest an impairment of attentional competency and interaction of the avatar. Another way of depicting relational attributes may be afforded by slowing down, distorting or dynamic compression of shouting originating from opponents and thus emphasising the avatars' perception of invested energy into a confrontation. In the same regard dynamics and expressive content in music influence how players relate differentially to fictional characters. These effects can be mapped onto relational attributes that span from agents corresponding to the subject [intro-], object [intra] and environment [interrelation] (cf. Gasselseder 2012, Gasselseder & Dobler 2009, Wingstedt et al. 2008, Hoeckner et al. 2011).

In Gasselseder (2012) subjects rated relational attributes between extro-avatar characters [= all characters except the avatar] differently depending on the music an action-adventure video game was accompanied with. For interrelational attributes, the overall experience of character interactions was rated as being more 'multifaceted' when playing the game in its dynamic music as compared to its non-dynamic music modality. For intrarelation, 'credibility' was linked to responses on items referring to perceiving characters from the perspective of the avatar. Higher ratings on credibility were seen when gameplay was accompanied by dynamic music. Introrelational attributes, however, seem less directly affected by music modality, but more sensitive towards sensory-spatial components that link embodied communication between the avatar and the user. Gasselseder and Dobler (2009) report a series of focus interviews where the presentation of sound-fx [as compared to music-only conditions] resulted in more frequent positive attributions of competency in avatar-user interactions following gameplay. This result goes in line with a previous study by Grimshaw and co-workers (2008) that found sound-fx to support the experience of flow; a mental state of absorption/involvement in performing actions. Thus, there is reason to assume that sound-fx may primarily take effect on introrelational attributes that negotiate syntax of embodied interactions. Further applications, then, could take advantage of personality traits or real-time psychophysiological measures of the player, so as to modulate expressive fidelity in a sense of sonic rules of interaction according to experienced emotional arousal. Another challenge in investigating the application of expressive fidelity is cross modal coherence as illustrated in subjects underestimation of distance and room size by giving lower presence ratings on visually corresponding acoustic room models as opposed to non-corresponding room acoustics (cf. Larsson et al. 2010; Knapp & Lomis 2004). While interacting with the external world may shape indeed our experience of expression, proto-physical conceptions of perspective may interfere when applying cross-modal realistic criteria on designing sound stimuli for immersive interactive media. In the case of the critically acclaimed side-scrolling game 'Limbo' (Playdead 2010) a reversed approach is pursued by presenting the player a 2D-canvas while sound is provided from an immersive first person perspective. Thus, matching cross-modal physical properties [as opposed to the afore mentioned case of perspective] may be less reliable in linking expressive cues within a situational context in contrast to the syntax exhibited in objects and environments.

6. CONCLUDING REMARKS

Whereas the present paper has given a schematic outline on how a recordings' artifact of expressive fidelity may be operationalised in experiencing multimedia, many questions remain on how expressive cues are integrated in situational contexts as well as how they evolve as conventions. A first account on the latter question may be seen in a study by Wingstedt and colleagues (2008) where the most dominate effects of music expression corresponding to a scene were observed in the reverberation level dialled in by subjects in addition to controlling for expressive parameters such as instrumentation. While these results intuitively exemplify the multifaceted relevance of the concept expressive fidelity in function to static scenes, further inquiries will be necessary in order to obtain a more detailed understanding on the acquired syntax of sonic expression in interactive media. Within the scope of audio in video games, a set of acoustic characteristics related to expressive congruency has been found to affect gameplay experience and performance, in particular when task goal definitions of gameplay are controlled for (cf. Gasselseder 2012). Following this, an on going empirical investigation currently being carried out by the author will provide a closer look at how properties like frequency bandwidth, dynamic modulation or background masking affect particular performance and experience criteria, as for instance the diversity of user action and the perceived quality of interface design. In addition to these points, the notion of expressive fidelity promises to demonstrate interpersonal differences in functional sound perception and is expected to find its way into recommendations for optimising ludosonic interaction design.

REFERENCES

Allesch, C. G. & Krakauer, P. M. (2006). Understanding our experience of music. What kind of psychology do we need? *Musicae Scientiae*, *10*, 1, pp. 41-63.

Baron-Cohen, S. (2003). *The essential difference*. Male and female brains and the truth about autism. New York: Basic Books.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts. Manifestations of a compositional system of perceptual symbols. In Collins, A. F., Gathercole S. E., Conway, M. A. & Morris, P. E. (Eds): *Theories of memory*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 29-101.

Begault, D. R., Wenzel, E. M. & Anderson, M. R. (2001). Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, *49*, 10, pp. 904-916.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton Century Crofts.

Blesser, B. & Salter, L.-R. (2007). *Spaces Speak, Are You Listening?* Experiencing Aural Architecture. Cambridge, MA: MIT Press, p. 361.

Boltz, M. G., Ebendorf, B. & Field, B. (2009). Audiovisual interactions. The impact of visual information on music perception and memory. *Music Perception*, *27*, 1, pp. 43-59.

Brenton, H., Gillies, M., Ballin, D. & Chatting, D. (2005). The Uncanny Valley. *Procedures of 19th British HCI Group Annual Conference*. Workshop on Human-Animated Character Interaction, Edinburgh, UK.

Bresin, R. & Friberg, A. (2011). Emotion rendering in music: Range and characteristic values of seven musical variables. *Cortex*, *47*, 9, pp. 1068-1081.

Bruner, J. (1986). Actual Minds, Possible Worlds. Harvard University Press, Cambridge, MA.

Bruner, J. S. & Postman, L. (1949). On the perception of incongruity. A paradigm. *Journal of Personality*, 18, pp. 206-223.

Chueng, P. & Marsden, P. (2002). Designing Auditory Spaces to Support Sense of Place. The Role of Expectation. Position paper for The Role of Place in On-line Communities Workshop, CSCW2002, New Orleans, November 2002. http://scom.hud.ac.uk/scompc2/research.htm (retrieved 30.01. 2015).

Eschrich, S., Münte, T. F. & Altenmüller, E. O. (2008). Unforgettable film music. The role of emotion in episodic long-term memory for music. *BMC Neuroscience*, *9*, 48, pp. 1-7.

Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses. *Frontiers in Psychology*, *3*, 151, pp. 1-20.

Gasselseder, H.-P. (2012). *Re-Orchestrating Game Drama*. The Immersive Experience of Dynamic Music in Video Games. Diploma thesis at the Department of Psychology, University of Salzburg, Salzburg: Austria, 349 pp. Due to impending publication, available online for a limited amount of time at https://www.box.com/s/zv2pxksmmd4e8knhyduv (retrieved 15.01. 2015).

Gasselseder, H.-P. & Dobler, V. (2009). *SPIELfilmTon* [GAMES-film-Sound]. Unpublished research paper at the University of Salzburg.

Gaver, W. W. & Mandler, G. (1987). Play it again, Sam. On liking music. *Cognition and Emotion*, 1, pp. 259-282.

Gray, K. & Wegner, D. M. (2010). Blaming God for Our Pain. Human Suffering and the Divine Mind. Personality and Social Psychology Review, 14, 1, pp. 7-16.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *290*, 1038, pp. 181-197.

Grimshaw, M., Lindley, C.A. & Nacke, L. (2008). Sound and Immersion in the First-Person Shooter. Mixed Measurement of the Player's Sonic Experience. *Development*, 1-7. http://www.bth.se/fou/forskinfo.nsf/17e96a0dab8ab6a1c1257457004d59ab/d48c2010f081aea6c1257521007f7924!OpenDocument (retrieved 13.05. 2015).

Hoeckner, B., Wyatt, E. W., Decety, J. & Nusbaum, H. (2011). Film Music Influences How Viewers Relate to Movie Characters. *Psychology of Aesthetics, Creativity, and the Arts*, *5*, 2, pp. 146-153.

Irrational Games (Developer) (2013). *Bioshock Infinite*. Microsoft Windows, Playstation 3, Xbox 360. Novato, CA: 2K Games.

Janata, P. (2009). Music and the Self. In Haas, R. & Brandes, V. (Eds.): *Music that works*. Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology. Vienna: Springer.

Jauk, W. (2007). *Music - a role model for a theory of media-arts.* Brief statement / Poster presentation at re:place, Berlin. http://www.uni-graz.at/~jauk/site.php?show=52 (retrieved 30.01. 2015).

Jauk, W. (2012). Medien der Erkenntnis und Gestaltung von Erlebniswelten zwischen embodiments und Metaphern [Media of cognisance and design of worlds of experience between embodiments and metaphors]. In Enders, B. (Ed): *Die Metapher als Medium des Musikverstehens*. Osnabrücker Beiträge zur Systematischen Musikwissenschaft, pp. 173-186.

Juslin, P. N. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance. Different channels, same code? *Psychological Bulletin*, 129, 5, pp. 770-814.

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G. & Silva, A. (2008). An experience sampling study of emotional reactions to music. Listener, music, and situation. *Emotion*, *8*, 5, pp. 668-683.

Kallionpää, M. E. & Gasselseder, H.-P. (2012). *Electro-Acoustic Music for Piano and Toy Piano. Music of Our Time?* Paper presented at the Reflective Conservatoire Conference. Performing at the Heart of Knowledge, Guildhall School of Music & Drama, London, GB, 18 March 2012.

Knapp, J. M. & Loomis, J. M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence, Teleoperators and Virtual Environments*, *13*, 5, pp. 572-577.

Kreutz, G., Schubert, E. & Mitchell, L. A. (2008). Cognitive Styles of Music Listening. *Music Perception*, *26*, 1, pp. 57-73.

Larsson, P., Väjamäe, A., Västfjäll, D., Tajadura-Jiménez, A. & Kleiner, M. (2010). Auditory induced presence in mediated environments and related technology. In Dubois, E., Gray, P. & Nigay, L. (Eds.): *The Engineering of Mixed Reality Systems*. London: Springer, pp. 143-163.

Leman, M. (2010). Music, Gesture, and the Formation of Embodied Meaning. In Godøy, R. I. & Leman, M. (Eds.): *Musical Gestures*. New York, NY: Routledge, pp. 126-153.

Nintendo (Developer) (1983). *Super Mario Bros.* Nintendo Entertainment System. J: Nintendo.

Ozawa, K., Ohtake, S., Suzuki, Y. & Sone, T. (2003). Effects of visual information on auditory presence. *Acoustical Science and Technology*, *24*, 2, pp. 97-99.

Playdead (Developer) (2010). *LIMBO.* Microsoft Windows, Apple OS X, Linux, Playstation 3, Xbox 360, iOS. Copenhagen, DK: Playdead.

Popper, A. N. & Fay, R. R. (1997). Evolution of the ear and hearing. Issues and questions. *Brain, Behavior and Evolution*, *50*, 4, pp. 213-220.

Ruby, P. & Decety, J. (2001). Effect of subjective perspective taking during simulation of action. A PET investigation of agency. *Nature neuroscience, 4,* 5, pp. 546-550.

Schubotz, R. I. (2007). Prediction of external events with our motor system. Towards a new framework. *Trends Cogn Sci, 1,* 5, pp. 211-218.

Slater, M. (2002). Presence and The Sixth Sense. *Presence.* Teleoperator's & Virtual Environments, 11, 4, pp. 435-439.

Wingstedt, J., Brändström, S. & Berg, J. (2008). Young adolescents' usage of narrative functions of media music by manipulation of musical expression. *Psychology of Music*, *36*, 2, pp. 193-214.

Zatorre R. J. & Halpern A. R. (2005). Mental concerts. Musical imagery and auditory cortex. *Neuron, 47*, 1, pp. 9-12.

Zwaan, R. A. (1999). Situation models. The mental leap into imagined worlds. *Current Directions in Psychological Science*, 8, pp. 15-18.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

USING THE MPPA ARCHITECTURE FOR UCT PARALLELIZATION

Aline Hufschmitt, Jean Mehat and Jean-Noël Vittaut
LIASD - University of Paris 8, France

ABSTRACT

We present here a study of the use of a *Multi-Purpose Processor Array* (MPPA) architecture for the parallelization of the UCT algorithm applied to the field of *General Game Playing*. We evaluate the constraints imposed by this architecture and show that the only parallelization of UCT proposed in the literature that is feasible on MPPA is a *leaf parallelization*. We show that the MPPA provides good scalability when increasing the size of the communications, which is useful when using synchronous communications to send large sets of game initial positions to be processed. We consider two approaches for the calculation of the playouts: the distributed computing of a playout on each cluster and the calculation of several playouts per cluster; we show that the second approach gives better results. Finally, we describe experiments concerning the thread management and present a surprising result: it is more efficient to revive threads than keep them alive and try to communicate with them.

KEYWORDS

Multi-Purpose Processor Array, Parallelization, UCT, MCTS, General Game Playing

1. INTRODUCTION

General Game Playing (GGP) is a branch of Artificial Intelligence with the aim of achieving versatile programs capable of playing any game without human intervention. These programs must be able to analyze the rules of an unknown game, to understand the goals, to discover the sequences of moves leading to victory and to play with expertise.

To find out what move to select at each stage of the game, different techniques of tree search have been developed to explore every position of the game and the different branches of possibilities which depend on the selected moves. The root of the tree represents the initial state of the game and the goal is to reach one of the leaves corresponding to the end of the game the score of which is the highest possible for the player.

Currently the algorithm used by most players is the *Upper Confidence bound applied to Trees* (UCT), a member of the *Monte Carlo Tree Search* (MCTS) algorithms family [Browne et al., 2012]. The number of game simulations (playouts) done by UCT determines the quality of the evaluation of a game tree node [Kocsis and Szepesvári, 2006]. Therefore, various approaches of UCT parallelization have been proposed to increase the number of playouts made in a given time.

Former GGP players used Prolog to interpret the language used to describe the rules, i.e. to determine the current state of the game, the list of legal moves or the scores in terminal states. These calculations are slow [Björnsson and Schiffel, 2013] but recently, the use of *Propositional Networks* (propnets) has allowed an important improvement in playout computation speed.

Different parallelization techniques of MCTS algorithms have been explored so far on machines with multi-core and/or multi-threads CPU used alone or networked to provide more computing power. Some of these techniques are applied to *General Game Playing* [Méhat and Cazenave, 2011a; Finnsson, 2012] but none, to our knowledge, realizes this parallelism on a player using a *propnet*. The important acceleration in playout computation brings new conditions for parallelization of UCT because communication and synchronization times become significant.

In this article, we study the parallelization of UCT using a *propnet* to interpret game rules on a *Multi-Purpose Processor Array* (MPPA), a new architecture marketed since 2013, created by the Kalray Company (Essonne, France) and dedicated to many-core processing. This very recent architecture has barely been tested on practical applications, thus we have chosen to explore the possibilities it can offer and evaluate its limitations.

Section 2 describes the different parallelization techniques proposed in the literature. Section 3 presents the MPPA architecture. Section 4 presents the feasible parallelization on MPPA. Section 5 evaluates the scalability enabled by this architecture for communications of varying size. Section 6 establishes the best approach to compute the playouts and section 7 compares different ways of managing threads inside clusters.

2. PARALLELIZATION APPROACHES OF UCT

Tree parallelization [Gelly et al., 2006] involves several processes in the construction of the game tree in a single shared memory. The use of a global mutex for the whole tree causes a bottleneck, so various improvements consisting in the use of local mutexes [Chaslot et al., 2008; Chaslot, 2010] or a lock free algorithm [Enzenberger and Müller, 2010] have been proposed.

On a distributed memory system, *root parallelization* [Cazenave and Jouandeau, 2007] consists in developing several distinct UCT trees from a position of the game. The different evaluations are collected after some time and combined to choose the best move according to different strategies (Best, Sum, Sum10, Rave, majority vote) [Méhat and Cazenave, 2011b; Soejima et al., 2010]. An improvement consists in synchronizing the evaluations of the top nodes at regular intervals during search [Gelly et al., 2008].

With *leaf parallelization* [Cazenave and Jouandeau, 2007], a single tree is constructed by a master process sending the positions to be explored to slave processes calculating the playouts. To minimize communication costs, Finnsson [2012] has proposed to realize multiple playouts per position, but some unworthy positions are then over-exploited. Chaslot et al. [2008] have proposed to stop a group of playouts that does not look promising to search again from another position but this approach does not provide good scalability. Cazenave and Jouandeau [2008] have suggested the use of asynchronous communications, which provides better scalability.

The *UCT-Treesplit* algorithm [Graf et al., 2011; Schaefers et al., 2011] combines the advantages of parallelization on distributed memory with the construction of a single tree. The machine on which the node is stored is selected using a hash key associated with the node. One drawback is the importance of communications necessary to perform a descent in the tree, i.e. select the start of a playout, or to update the evaluation of nodes. Yoshizoe et al. [2011] have proposed to solve this problem by performing a depth-first search.

3. MPPA ARCHITECTURE

The MPPA-256 chip is a multi-core processor composed of 256 processing cores (PC) organized in a grid of 16 clusters connected through a high-speed *Network-on-Chip* (NoC). It is the first member of the MPPA MANYCORE family, the others reaching up to 1024 processors in a single silicon chip.

The MPPA-256 is a MIMD architecture with a distributed memory accessible only locally, i.e. each cluster is encapsulated, only accessing its own memory, and can execute code independently of other clusters.

Each cluster contains 16 PC, a shared memory of 2MB and a system core using a specific operating system (NodeOS). This system core supervises the scheduling, execution of tasks and data transfers while the 16 PC are dedicated to application code.

Four I/O interfaces allow communications between the host machine and clusters for two of them and between clusters and the Ethernet network for the two others, but with the current version of the middleware we only have access to one of the I/O interfaces: the software tools designed for the MPPA and grouped under the name of MPPA ACCESSCORE are currently under development. This I/O interface has a quad core SMP processor with a 4GB DDR3 memory and a PCI Gen3 interface for communications with the host. Different connectors are available to implement synchronous or asynchronous communications using a simple buffer or queues.

A technical report Jouandeau [2013] states that "the computing capabilities of Intel i7 3820 processor with 8 cores and a MPPA processor with 256 cores are close". The estimate is based on the performance observed for a single core on solving the spin glass problem and multiplied by the number of cores, which implies parallel algorithms with zero communication times.

The limitation a cluster's memory to 2MB is a major constraint in the development of applications for the MPPA. In our case, we see in section 4 that it reduces the choices of possible parallelization approaches. Even if the MPPA-256 card seems to be limited in its performance, we must concede that it is only the first member of the MPPA MANYCORE family. A *Coolidge* processor with 1024 cores is expected in 2015. In addition, several MPPA-256 cards can be used together in the same host machine to increase computing capacity. Therefore, the MPPA architecture presents possibilities for the evolution of computing power that deserve investigation.

4. PARALLELIZATION OF UCT ON MPPA

Our experiments were carried out on a server equipped with an Intel Core I7 at 3.6GHz running Linux OS and a PCIe Application Board AB01 equipped with a chip MPPA-256 [Kalray, 2013]. Our program is based on Jean Noël Vittaut's *LeJoueur*, written in C++.

The limited size of memory inside the clusters does not allow the construction of a UCT tree. Therefore neither a *tree parallelization* inside each individual cluster nor a *root parallelization* nor a *UCT-Treesplit* can be performed on a MPPA. The only remaining choice is a *leaf parallelization*. Unfortunately, this technique is less effective than *root* or *tree parallelizations* because it suffers from limitations caused by the master process and communications [Soejima et al., 2010]. In addition to this, memory limitations imply a coding as concise as possible so that the *propnet* can fit into the cluster memory. For the most complex games, like Hex, the size of the *propnet* or the size of the positions to be processed exceeds the available space or connectors capacity.

For all the experiments presented in this article we used synchronous communications. This allows us to have benchmark results that we can later compare to the ones using asynchronous communications. Sets of game positions are sent to the MPPA from which playouts are computed. The I/O interface waits for the results from all the clusters before sending them back to the host. Each transmission of a position set, calculation of the playouts and recovery of the results (the scores) is referred to as a *run*.

5. SCALING FOR VARIABLE SIZE OF COMMUNICATIONS

In GGP, the size of position descriptions varies significantly from one game description to another in a ratio from one to ten: we need communications of about 200 to 2000 bytes to send a position to the MPPA. In the case of synchronous communications, if we want to make one playout per thread (i.e. per PC), the size of a communication is the size of a position multiplied by the number of threads.

To evaluate the scalability of the MPPA we varied the size of messages sent between the host and the I/O node and between the I/O node and the clusters from 200 to 2000 bytes. To prevent the variable time required to calculate the playouts from disturbing measurements, no playout was actually calculated and an empty result was returned immediately. Each execution of the program performs 1000 runs. We measured performance as the time necessary to execute these 1000 runs for 1 to 16 clusters with 1 to 16 threads per cluster.

Figure 1 presents the experimental results. With one thread on one cluster, we can see that the running time is almost constant while the size of the communications increases tenfold. With 16 threads on one cluster, we can see an additional time corresponding to the time required to start the threads and the time necessary to distribute the data among threads. The curve corresponding to 16 clusters with one thread each shows the time required for the I/O node to distribute data: the connector is set to the destination cluster before sending each part of the data.

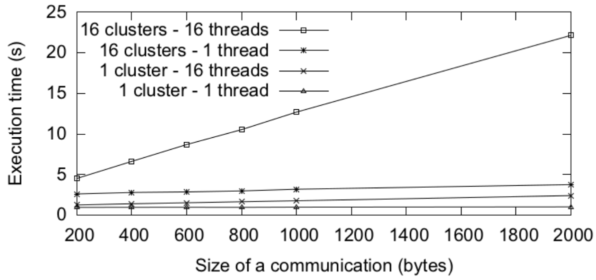


Figure 1. Evolution of the execution time depending on the size of communications (1000 runs)

Scaling is then constrained by two aspects: first, when the I/O node receives a set of positions it needs time to distribute them among the clusters and second, the thread start time inside a cluster is not negligible. Scaling is only slightly hindered when considering these two aspects separately but when we combine them (for 16 clusters and 16 threads per cluster), we find that it significantly degrades it. However, we see that scaling remains satisfactory since the execution time increases almost fivefold when the message size increases tenfold.

6. PLAYOUT CALCULATION APPROACHES

In the restrictive conditions of a *leaf parallelization*, the choice remains of the playout calculation approach. The *propnet* is represented by a logic circuit with different layers of logic gates. Each layer is represented by a set of rules, which can be evaluated in any order. We present here two approaches for the calculation of playouts with a *propnet*. In the first one, we distributed the calculation of each rule layer among the different threads of a cluster and then processed a single playout per cluster. In the second we proceeded to the calculation of a complete playout on each thread of a cluster, obtaining N playouts per cluster.

We conducted these experiments on three games: *Tictactoe* which has short playouts (between 5 and 9 moves) and a small quick to evaluate *propnet*, *EightPuzzle* which has a small *propnet* too but playouts up to 60 moves and *Breakthrough* which has a *propnet* which is longer to evaluate. Each execution of the program performs 10000 runs. We measured the performance based on the number of playouts carried out per second.

Figure 2 presents the performance obtained on the game of *Tictactoe* for 1, 2, 4, 8, 16 clusters with 1, 2, 4, 8, 16 threads per cluster and demonstrates the significant cost in synchronization generated by the division of playout computation. Each layer of the *propnet* circuit depends on its predecessor, so the threads must synchronize. The calculation of each part of a playout is considerably speeded up by the use of a *propnet*. The distributed computing of a playout cannot therefore provide any benefit considering the overhead introduced by the synchronization barrier between layers.

Figure 3 presents the results for the three games with 1 to 16 clusters and 1 to 16 threads per cluster. It shows that the performance scales well when a complete playout is computed on each PC.

Results for the game of *Tictactoe* show a progression from ≈450 playouts/s for one cluster with one thread to ≈77700 playouts/s for 16 clusters with 16 threads, i.e. a speed increase of 170 for a 256 PC machine.

For the game *EightPuzzle*, the speed increase is, at best, 133 (for 14 threads per cluster) with 224 PC. The performance does not scale well after 14 threads. We explain this result by the constant length of playouts: playing randomly, the solution has little chance of being found and playouts are stopped after 60 moves by the stepper. Therefore, all the scores are sent to the I/O node at the same time by all clusters. The size of the sent messages increases with the number of threads and the communications get slower. A similar deceleration hindering scalability can be observed in *Tictactoe* by forcing the players to completely fill the grid for each game: playout length is then set to 9 moves.

Scaling is better for *Breakthrough*. Computation time is longer, therefore communication and synchronization times are lower in comparison. We can observe a speed increase of 155 for a 256 PC machine.

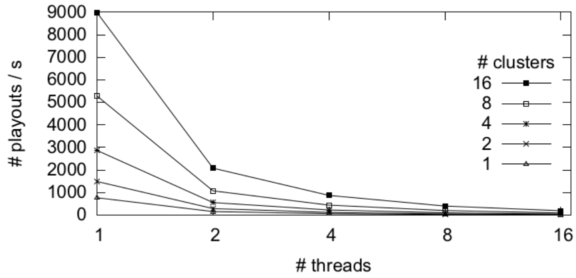


Figure 2. Evolution of the performance for the game of *Tictactoe* i.e. number of playouts per second (for 10000 runs) with calculation of a playout distributed on N threads

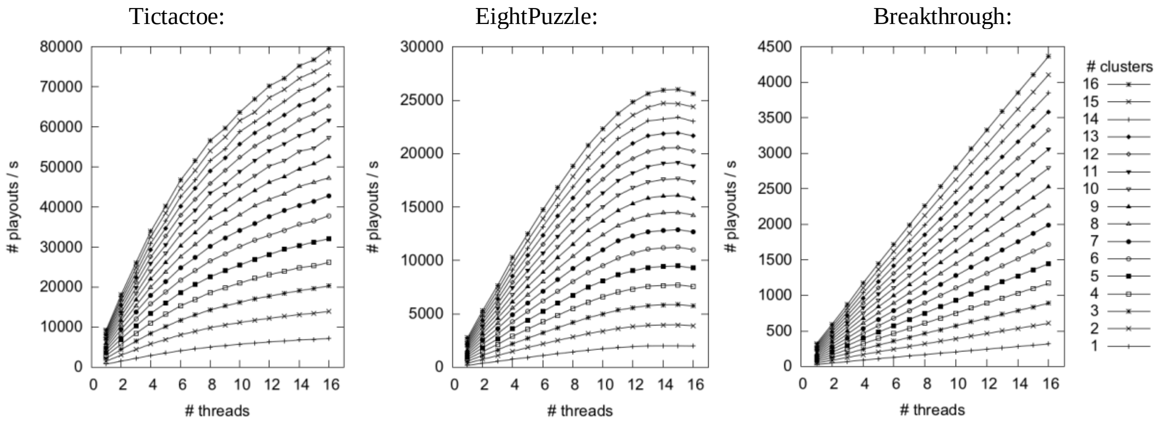


Figure 3. Evolution of the performance i.e. number of playouts per second (for 10000 runs) with the calculation of one playout per thread

7. THREAD MANAGEMENT

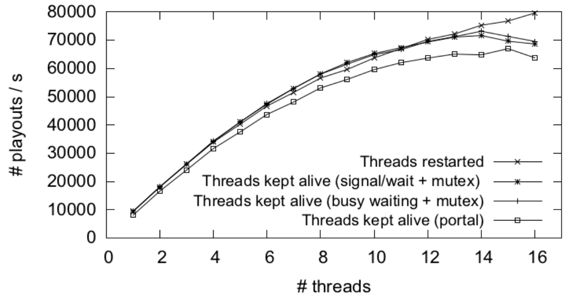
In the previous experiment threads were restarted for each set of playouts. Another strategy is to keep the threads alive waiting for playout requests. To test different approaches for communications between a cluster and its threads, we take the second experiment of the previous section and change the management of threads. Each cluster receives a set of positions it distributes to the different threads. The tests are performed on the different games by varying the number of clusters and threads per cluster; 10000 runs are executed every time.

In the first part of the experiment we use the Posix functions `pthread_cond_wait` and `pthread_cond_signal` to communicate with the threads. In the second part, we replace the Posix functions by a busy wait. In the third part we use portal connectors which are part of the MPPA ACCESCORE tools and uses the NoC to create communication paths between the main thread of the cluster and the other threads. This third approach avoids the use of mutexes.

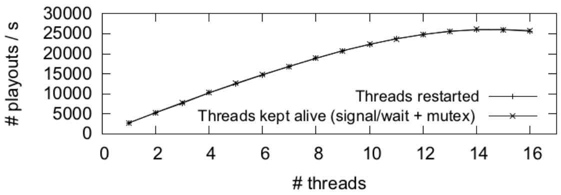
Results are displayed in figure 4 for 16 clusters and 1 to 16 threads per cluster and compared with results obtained with the second experiment of the previous section where threads were restarted. For the games *EightPuzzle* and *Breakthrough,* we compare only the restart of the threads with the use of the Posix functions `pthread_cond_wait` and `pthread_cond_signal`.

With the game of *Tictactoe*, the use of the portal is less effective regardless of the number of threads used, and the different approaches keeping threads alive does not scale well above 14 threads per cluster. We note with surprise that the technique consisting in restarting threads for each request of calculation provides a better scalability and gives the best results. The results obtained on the games *EightPuzzle* and *Breakthrough* do not show such a marked difference since the curves are overlaid, but confirm that keeping threads alive provides no benefit.

Tictactoe:



EightPuzzle:



Breakthrough:

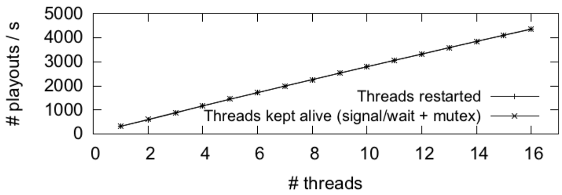


Figure 4. Evolution of the performance i.e. number of playouts per second (10000 runs, 16 clusters) with different thread management modes

This can be explained by the use of a shared memory inside each cluster and the fact that all the synchronization primitives end on a spinlock or equivalent. When a thread has finished its work, it waits for a signal by polling in the shared memory. Therefore, if the other threads have not finished the calculation of their playout, they are slowed down since they also need to access the shared memory. The main process running on PC0 and responsible for the communications with the I/O node is also slowed down. The more threads finish their task, the greater the execution of another process is slowed. On the contrary, when a thread uses `pthread_exit`, the PC running this thread is placed in an idle state, so it does not disturb the work of other PCs. Halting and relaunching the threads is then more efficient.

8. CONCLUSION

In this paper we have studied the capabilities offered by a *Multi Purpose Processor Array* (MPPA) architecture for the parallelization of the UCT algorithm in the field of *General Game Playing*. The limitation of the memory inside the cluster to a size of 2MB is the major constraint. Among various parallelization techniques described in the literature the only applicable one, with this limited memory space, is *leaf parallelization*.

We have demonstrated that the MPPA provides good scalability when increasing the size of communications, giving good results when using synchronous communications and sending large sets of game initial positions to be processed in a single run.

We have considered two approaches for the calculation of playouts. The distributed computing of a playout on the different PC of a cluster causes an important synchronization overhead. The calculation of a complete playout per PC gives better results.

We were able to establish that on MPPA it is more efficient to restart threads for each calculation request. All synchronization primitives on the MPPA end on a spinlock. Therefore, threads kept alive slow down the working ones because of memory access competition.

It would be possible to reduce the weight of communications by making several playouts from each sent position, but carrying out several playouts from the same position may be less beneficial for the UCT exploration than starting from different positions [Chaslot et al., 2008].

One may think that the use of asynchronous communications would improve these results but unfortunately the first experiments we have conducted have yielded disastrous results. This comes from middleware problems that our experiments have revealed. These problems should be fixed by Kalray in the next release. The use of asynchronous communications will therefore be the subject of future experiments.

We should also compare our results with what could be achieved with a GPU using a framework like Cuda on the same problem. The SIMT architecture requires the use of synchronous operations but this can be an advantage to distribute the calculation of each layer of the *propnet* without the need for a specific synchronization barrier. Moreover, the large quantity of memory shared by computing units allows the implementation of the different parallelization techniques presented.

Future works also include the test of new approaches for UCT parallelization or modifications of UCT. For example, the creation of mini-UCT-trees in cluster memory can save communication costs. The SHOT alternative to UCT [Cazenave, 2015] offers interesting perspectives as it scales well in addition to using less memory and it can be efficiently parallelized.

REFERENCES

Björnsson,Y.and Schiffel, S. 2013. Comparison of gdl reasoners.In *Proceedings of the IJCAI-13 Workshop on General Game Playing (GIGA'13)*, pp. 55–62.

Browne, C. B., et al. 2012. A survey of monte carlo tree search methods. *Computational Intelligence and AI in Games, IEEE Transactions on*, 4(1):1–43.

Cazenave, T. 2015. Sequential Halving Applied to Trees. IEEE Trans. Comput. Intellig. and AI in Games 7(1): 102-105

Cazenave, T. and Jouandeau, N. 2007. On the parallelization of UCT. In *Proceedings of the Computer Games Workshop*, pp. 93–101.

Cazenave, T. and Jouandeau, N. 2008. A parallel monte-carlo tree search algorithm. In van den Herik, H. J., Xu, X., Ma, Z., and Winands, M. H. M., editors, *Computers and Games*, volume 5131 of *Lecture Notes in Computer Science*, pp. 72–80. Springer.

Chaslot, G. 2010. *Monte-Carlo Tree Search*. PhD thesis, Universiteit Maastricht.

Chaslot, G., Winands, M. H. M., and van den Herik, H. J. 2008. Parallel monte-carlo tree search. In *Computers and Games*, pp. 60–71.

Enzenberger, M. and Müller, M. 2010. A lock-free multithreaded monte-carlo tree search algorithm. In *Proceedings of the 12th International Conference on Advances in Computer Games*, ACG'09, pp. 14–20, Berlin, Heidelberg. Springer-Verlag.

Finnsson, H. 2012. *Simulation-Based General Game Playing*. Doctor of philosophy, School of Computer Science, Reykjav ́ık University.

Gelly, S., et al. 2008. The parallelization of monte-carlo planning - parallelization of mc-planning. In Filipe, J., Andrade-Cetto, J., and Ferrier, J.-L., editors, *ICINCO-ICSO*, pp. 244–249. INSTICC Press.

Gelly, S., et al. 2006. Modification of uct with patterns in monte-carlo go. Technical Report 6062, Inria.

Graf, T., et al. 2011 .Parallel monte-carlo tree search for hpc systems. In *Proceedings of the 17th International Conference on Parallel Processing - Volume Part II*, Euro-Par'11, pp. 365–376, Berlin, Heidelberg. Springer-Verlag.

Jouandeau, N. 2013. Intel versus mppa. Technical report, LIASD Universit Paris8.

Kalray 2013. *MPPA ACCESSCORE 1.0.1 - POSIX Programming Reference Manual - KETD-325 W08*. Kalray SA.

Kocsis, L. and Szepesvári, C. 2006. Bandit based monte-carlo planning. In *Proceedings of the 17th European Conference on Machine Learning*, ECML'06, pp. 282–293, Berlin, Heidelberg. Springer-Verlag.

Méhat, J. and Cazenave, T. 2011a. A parallel general game player. *KI*, 25(1):43– 47.

Méhat, J. and Cazenave, T. 2011b. Tree parallelization of ary on a cluster. In *GIGA 2011, IJCAI 2011*, Barcelona.

Schaefers, L., Platzner, M., and Lorenz, U. 2011. Uct-treesplit - parallel mcts on distributed memory. In *MCTS Workshop*, Freiburg, Germany.

Soejima, Y., Kishimoto, A., and Watanabe, O. 2010. Evaluating root parallelization in go. *IEEE Trans. Comput. Intellig. and AI in Games*, 2(4):278–287.

Yoshizoe, K., et al. 2011. Scalable distributed monte-carlo tree search. In *SOCS*.

ISBN: 978-989-8533-38-8 © 2015

TOPOLOGICAL CLUES FOR PREDICTING OUTCOMES OF MULTIPLAYER ONLINE BATTLE ARENA GAMES

Dominik Deja[1] and Mateusz Myślak[2]
[1]*Polish-Japanese Academy of Information Technology*
[2]*Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics*

ABSTRACT

With 27 million people playing *League of Legends* every day, e-sports became more and more important part of our everyday life. Rise of its popularity builds a demand for better understanding e-sports mechanics on a deeper level. In the article, we test a hypothesis that it is possible to predict an outcome of Multiplayer Online Battle Arena game based on topological clues only (such as area of polygon where vertices are players positions, and their dispersion). With accuracy, precision, and recall around 80%, we show that it is possible to construct a model which will be both: generic for MOBA games genre, and based on simple assumptions, which are hard to be perceived by the naked eye. Results can be used for better understanding how MOBA games works, as well as for developing more player-friendly games in terms of difficulty, and time one needs to learn how to play.

KEYWORDS

Esports, MOBA, Predictive Analytics, Data Mining

1. INTRODUCTION

Up to now, sports data mining became a fairly well ingrained in trainers and sport managers consciousness (Schumaker et al. 2010). Most of the best teams in football, basketball, and baseball are backed up by data scientists, who continuously track team's behavior in search for hints, which would leverage their effectiveness. Yet, from a research point of view, one of the disadvantages of those sports is scarcity of data. Thankfully, this is not the case with e-sports, one of the youngest branch of sport, which now develops in an astonishing pace (Taylor 2012).

The most popular online game, which is an important part of e-sports, *League of Legends* (*LOL*) gathers daily 27 million of people, with concurrent peak on 7.5 million of players. Every month it is played by 67 million of people[1]. Those numbers show that e-sports became an important part of our everyday life, and as such, deserves to be studied in detail.

In the article we use data from *LOL* in order to validate the hypothesis, whether topological clues derived from a game can be used to predict a game's outcome. As we show, one does not need to understand games mechanics; instead information about which side will win, can be predicted strictly from their behavior clues, which even for a trained eye might be hard to catch, yet which can be easily obtained.

For data acquisition we used a custom application written in C# supported with MSSQL database. For data preprocessing we used Python 2.7. For modeling we used the R language with its indispensable "data.table" (for operations on really big tables), and "randomForest" (for convenient modeling of random forests) packages.

[1] http://www.forbes.com/sites/insertcoin/2014/01/27/riots-league-of-legends-revealsastonishing-27-million-daily-players-67-million-monthly/

2. RELATED WORKS

Professionalization of e-sports is a fact. Taylor (2012) provides an exhaustive historical introduction to this process, and shows how e-games industry gained (and it is still gaining) its importance in the modern world. Bornemark (2013) examines what differentiates successful, in terms of popularity, multiplayer video games from failed ones. One of his findings is that the most popular games are the ones, which are supported by e-sports proactive developers. De Prato et al. (2014) study shows how the global market for video games evolved, and how rise in popularity of multiplayer games reshaped e-games industry. A number of case studies has been made: Carter & Gibbs (2013) analyze *EVE Online*, Rambusch et al. (2007) explore *Counter-Strike*, Taylor (2011) focus on *Halo 3*. Finally, Jonasson & Thiborg (2010) try to foresee the future of e-sports and its impact of sports in general.

Using data mining can clearly leverage teams' effectiveness. Thus, sports data mining is an area of broad research. Schumaker et al. (2010) is a considerable monograph in the field describing most of state-of-the-art solutions. Large part of the research focus on traditional sports, yet a number of articles treat about e-sports. Bosc et al. (2013) uses patterns discovery techniques in order to analyze predominant strategies in *Starcraft II*. Yang & Roberts (2013) also use knowledge discovery algorithms, but they expand their focus to *DotA*, *Warcraft III*, and *Starcraft II*. Medler (2012) elaborates on how e-game analytics influences players' experience.

For the best of our knowledge, analyzing topological clues for e-games has not been deeply studied. Kang et al. (2013) uses trajectory data from *World of Warcraft* in order to analyze players' behavior. Rioult et al. (2014) examine basic topological clues from a small sample of *DotA* games and on that base build a model for predicting a game's outcome. Yet, due to a small size of sample and a large bias in data they use, their findings cannot be perceived as fully convincing. Therefore, we would like to close this gap with our research, which overcame difficulties meet by our predecessors.

3. MULTIPLAYER ONLINE BATTLE ARENA GAMES

Even though Multiplayer Online Battle Arena games' genre (MOBA) basically started from being a group of home-made modifications for big and popular games such as *Starcraft*, or *World of Warcraft*[2], it has quickly became as hugely popular as its original ancestors. In late 2012, Forbes announced, that *LOL* is "the most played PC game in the world"[3]. As the game was created in 2009, it shows how big the demand for professionally developed MOBA game was.

Basic mechanics of MOBA games are simple: key objective for a player is to, along with his/hers four teammates, beat enemy team and destroy their base. In order to obtain this goal, each player controls one champion and has to use resources available on the map (mostly, these is gold obtainable via killing monsters, and enemies) to build as strong character as it is possible, and cooperate with his/hers own team. From this point, things get more complicated, as there are many ways how to do it, and variety of strategies and tactics can be applied.

---

[2] http://www.polygon.com/2013/9/2/4672920/moba-dota-arts-a-brief-introduction-togamings-biggest-most
[3] http://www.forbes.com/sites/johngaudiosi/2012/07/11/riot-games-league-of-legendsofficially-becomes-most-played-pc-game-in-the-world/

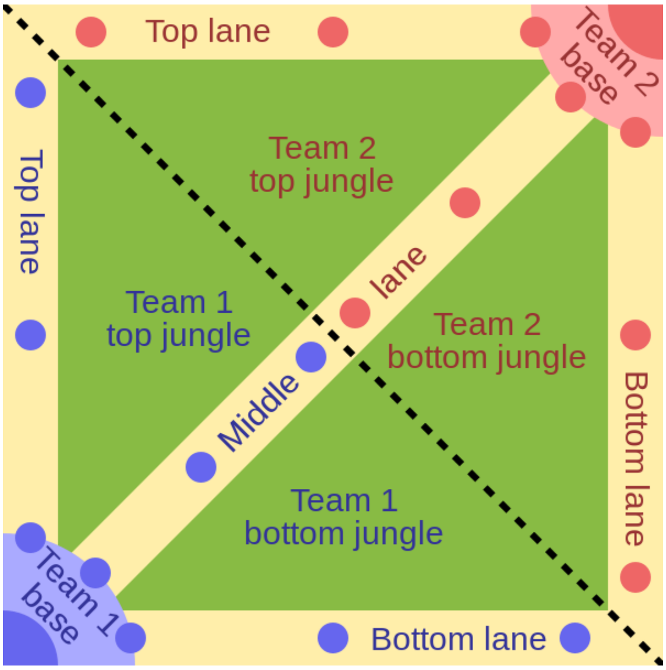


Figure 1. Scheme of MOBA game map (source: Wikipedia)

Focusing on game's topology, which is fairly similar to all MOBA games, a game's map is a square, divided into two sides according to top-left, bottom-right axis. Blue team has got their base in the bottom-left side of a map, while red team in the top-right corner. Connection between two bases is divided into three lanes, top, mid, and bottom lane, however there is also a jungle between them. A generic map is provided in Figure 1.

As a constant number of allied monsters, called ”minions”, or ”creeps”, leave base every fixed period of time, and follows lanes in order to seek and destroy enemy's base, it is reasonable for every team to locate at least one of their players on every lane, and leave one for jungle, where neutral monsters linger. By such allocation, players can effectively use resources available on a map, while keeping pressure on an enemy team.

League of Legends is a major title in MOBA games' genre. Thus, it effectively sets a standard for other games in industry. However, it is not free from minor deficiencies. One of them, is a slight unbalance of a map, which results in 51.98% chances of winning for team starting on a blue side. As LOL implements a rather solid matchmaking system which focuses on selecting balanced teams with each having equal chances of winning (Myślak & Deja 2014), this can only be seen as an unbalanced map, which favors one of the sides.

4. PREDICTING OUTCOME USING TOPOLOGICAL CLUES

Our goal was to build a model which predicts a game's outcome and is based solely on topological clues, which can be obtained from the game. To make our goal more ambitious, we confined ourselves to variables which does not lead to straightforward prediction of a match's outcome, and which does not requires deeper knowledge on game mechanics. By this, on the one hand we make our model more generic, while on the other we avoid the case where we try to predict outcome variable on a base of the highly correlated variables.

Example of the latter is a variable called "remoteness", used by Rioult et al. (2014). It is a distance from a team's actual position to their base location. It is obvious for every observer that since the goal of a game is to push back enemy team, and to destroy their base, there is almost a linear relationship between chances of winning a game, and value of remoteness. The farther team is able to push, the stronger it is. Obviously, using this variable in model boosts effectiveness almost to Bayesian error, but such model has literally no practical usage. Table 1 describes in details variables we used in a final model.

Table 1. Variables used in modeling

Variable Name	Description
AveTier	Mean of players' tiers (skill estimate)
WardsPlaced	How many wards were placed by a team members
WardsKilled	Percentage of destroyed enemy's wards
Area	Area of the convex polygon made out of players' positions
Gathering Ability	Mean players' distance from the barycenter of a convex polygon made out of their positions
Inertia	Standard deviation of players' distance from the barycenter of a convex polygon made out of their positions
Diameter	The biggest distance between players in a team

5. RESULTS

League of Legends data is freely available, and one may access it via LOL's API [4], thus our dataset was obtained from it. LOL's API provides an enormous amount of data, describing almost every step player takes in a game. From what he/she buys/sells at the shop, through champion's skin and selected skills, to a number of minions killed by a player every minute.

In total, we've collected 5 925 698 games from Season 5, out of which we selected 5 199 542 games for which we knew what is the approximation of all players' skills (also obtained from LOL's API), since LOL's shows player's estimated skill only if a player played at least 10 ranked games. Also we restrained ourselves to analyze only ranked 5vs5 games. Additionally, this restriction comes from the fact that ranked games tend to be more competitive than normal games (where one does not know what is his/hers skill estimate). Thus, we assumed that players tend to play up to their best and toxic behavior is less common. All this was an effort to reduce the level of noise in data.

Data acquisition was made as follows:

1. Obtain first ranked 5vs5 game played in Season 5 called $seed_0$.
2. From $seed_0$ find next n ranked 5v5 games - *seeds*.
3. For each *seed* in **seeds**, where **seeds** = [$seed_0,...,seed_n$]:

   For each *participant* in **participants**, where **participants** = [$participant_0,...,participant_{10}$]: Download all games played by *participant*.

Using this procedure, we obtained games from a period of 21 January 2015 09:30:02 GMT to 25 February 2015 15:26:54 GMT, while seeds only are from period of 21 January 2015 09:30:02 GMT to 23 January 2015 11:55:36 GMT. In total, we collected 438 285 seeds, out of which we collected almost 6 million games. It can be estimated, that since games were played in a 2.41 game/second pace, we obtained around 82% of all games played in this one month period. Due to the use of this procedure we obtained bias free dataset.

Information about tempo-spacial behavior of players is kept mostly as snapshots taken every minute. Every snapshot archives a number of features, including player's x, and y coordinates. To reduce variability of features derived from such data, we used aggregates. For every 5 minutes period in game, a mean of every variable was taken.

On average, as it is showed in Figure 2, game duration is around 35 minutes. Therefore, even though we computed all the 5 minutes lasting periods for every game, for final model, we used 8 periods for each variable (except for *WardsPlaced* and *WardsKilled*): [(0,5),(5,10)(10,15)(15,20)(20,25)(25,30)(30,35)(35,40)].

[4] https://developer.riotgames.com/

ISBN: 978-989-8533-38-8 © 2015

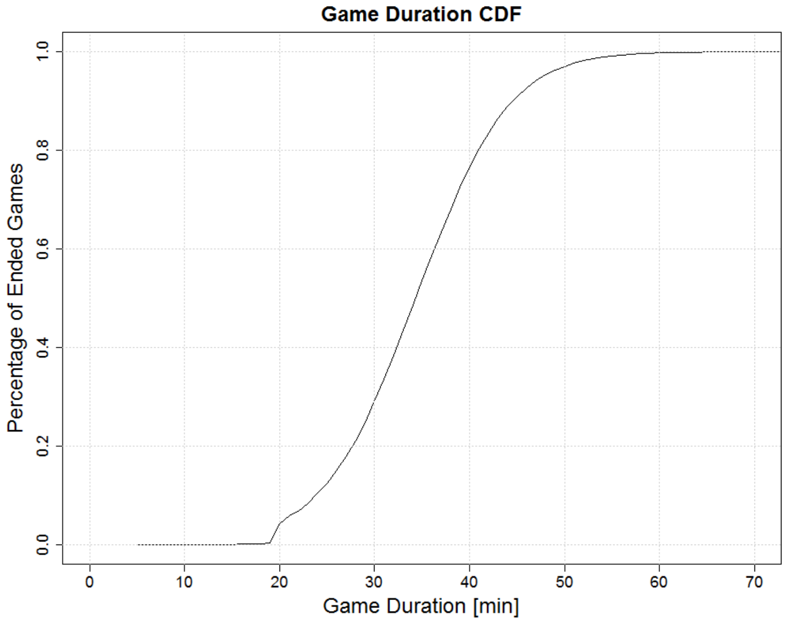


Figure 2. Cumulative Distribution Function of Game Duration

For modeling purposes, we decided to use random forests. Amid many state-of-the-art techniques they are known to possess one of the best effectiveness and robustness to all kinds of noise (Fernandez-Delgado et al. 2014). We run the model two times, with 1000 trees each. First time, we used all of the variables, at a second time we cut off variables with the least importance. Table 2 shows confusion matrix, while Table 3 illustrates standard statistics used to compare results of classification problems.

Table 2. Confusion matrix of a model

		True	
		Win	Lose
Predicted	Win	2136445	580202
	Lose	566246	1916648

Table 3. Evaluation estimates

Precision	Recall	F Measure	Accuracy
78.64%	79.05%	78.85%	79.95%

It is noticeable that Yang & Roberts (2013) using different kind of information achieved 83.5% accuracy (analyzing the data coming from DotA). Using “remoteness” variable we managed to obtain even better results: F Measure equal to 92.31%, and accuracy equal to 92.01%, which surpass results given by Rioult et al. (2014) – F Measure equal to 87.43%. Variables of the biggest importance (evaluated by mean decrease of Gini Index), are showed in Table 4.

Table 4. The most important variables

Variable	Mean Decrease of Gini Index
RGathering30	1613.154009
BGathering30	1433.803814
RGathering25	1353.38887
RDiameter30	1281.908946
BGathering25	1250.901231
RGathering35	1216.0548
RDiameter25	1142.821707
BDiameter30	1128.327775
BGathering35	1027.524783
RArea30	1026.47076

“B” denotes blue team, where “R” denotes red team. Therefore “RArea30” denotes mean area of red team’s polygon from 25 to 30 minute of a game. Interestingly, almost all the most important variables are the ones which relate to the distance between players and a barycenter of a convex polygon determined by the outer players.

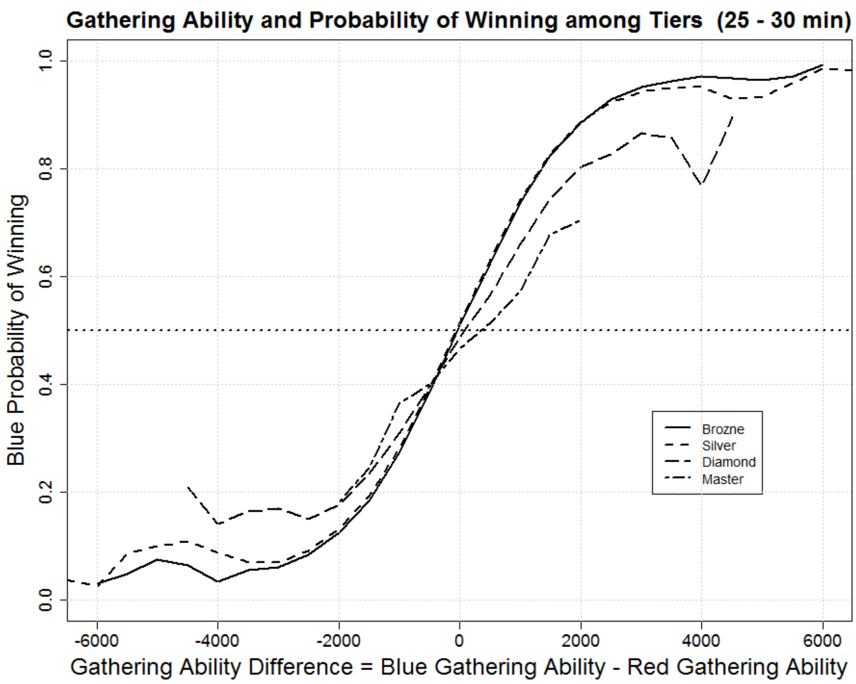


Figure 3. Together or in a distance?

Figure 3 shows that in period 25-30 minutes of a game, chances of winning increase along with the dispersion of team members. This behavior is called “split pushing”, and is a common way how to ensure one’s advantage in a game. Therefore model confirms that in order to win, a team has to maintain a broader map control than their enemies. As there are the two most important variables, it seems fairly reasonable to state that a map control is a key factor in winning a game. The fact that for lower triers cases more extreme are observed is due to higher variance in skill amid players in lower tiers. Higher ranks are more unified in terms of their abilities, and as such, they play rather equal (one has to keep in mind, that plot shows the relation between variables from two teams).

6. CONCLUSIONS

Results show that it is possible to make a highly accurate prediction of MOBA game's outcome solely from topological clues. This phenomenon can be interpreted as a consequence of a fact that MOBA games are highly based on teamplay. If players fail in forming a team, then their chances of winning a game drops drastically. If players succeeds, it means that they were more effective in using strategic resources and positioning on a map. This is what in a nutshell was tested by our research.

This knowledge may be leveraged by industry, in order to provide players with more information how to play in a way which will boost their effectiveness. It is useful for professional players who might overlook simple drawbacks in their play, as well as amateurs who wish to improve their skill.

In our future work we would like to focus on answering multiple questions consisting predictability of e-sports, as well as on how such information might be useful for learning how to play. Even though, basic mechanics of MOBA games are fairly simple, archiving better results requires deeper knowledge and greater skills. We would like to focus on how players can effectively learn how to play.

REFERENCES

Bornemark, O. (2013), 'Success factors for e-sport games', *USCCS 2013* p. 1.

Bosc, G., Kaytoue, M., Raıssi, C. & Boulicaut, J.-F. (2013), 'Strategic pattern discovery in rts-games for e-sport with sequential pattern mining'.

Carter, M. & Gibbs, M. R. (2013), esports in eve online: Skullduggery, fair play and acceptability in an unbounded competition., *in* 'FDG', pp. 47–54.

De Prato, G., Feijoo, C. & SIMON, J. P. (2014), 'Innovations in the video game industry: Changing global markets', *Communications & Strategies* (94), 17.

Fernandez-Delgado, M., Cernadas, E., Barro, S. & Amorim, D. (2014), 'Do we need hundreds of classifiers to solve real world classification problems?', *The Journal of Machine Learning Research* 15(1), 3133–3181.

Jonasson, K. & Thiborg, J. (2010), 'Electronic sport and its impact on future sport', *Sport in Society* 13(2), 287–299.

Kang, S.-J., Kim, Y. B., Park, T. & Kim, C.-H. (2013), 'Automatic player behavior analysis system using trajectory data in a massive multiplayer online game', *Multimedia tools and applications* 66(3), 383–404.

Medler, B. (2012), 'Play with data-an exploration of play analytics and its effect on player expereinces'.

Myślak, M. & Deja, D. (2014), Developing game-structure sensitive matchmaking system for massive-multiplayer online games, *in* 'Social Informatics', Springer, pp. 200–208.

Rambusch, J., Jakobsson, P. & Purgman, D. (2007), Exploring e-sports: A case study of game play in counter-strike, *in* 'Situated play: The 2007 world conference of Digital Games Research Association', DiGRA, pp. 157–164.

Rioult, F., Metivier, J.-P., Helleu, B., Scelles, N. & Durand, C. (2014), 'Mining tracks of competitive video games', *AASRI Procedia* 8, 82–87.

Schumaker, R. P., Solieman, O. K. & Chen, H. (2010), *Predictive modeling for sports and gaming*, Springer.

Taylor, N. T. (2011), 'Play globally, act locally: The standardization of pro halo 3 gaming', *International Journal of Gender, Science and Technology* 3(1).

Taylor, T. (2012), *Raising the Stakes: E-sports and the Professionalization of Computer Gaming*, Mit Press.

Yang, P. & Roberts, D. L. (2013), Knowledge discovery for characterizing team success or failure in (a) rts games, *in* 'Computational Intelligence in Games (CIG), 2013 IEEE Conference on', IEEE, pp. 1–8.

PERCEPTIONS OF THE ELDERLY USERS OF MOTION TRACKING EXERGAMES

Tapani N. Liukkonen[1], Tuomas Mäkilä[1], Hanna Ahtosalo[2], Toni Heinonen[2],
Reetta Raitoharju[3] and Paula Pitkäkangas[4]

[1]Technology Research Center, University of Turku, 20014 Turun yliopisto, Finland
[2]Turku Game Lab, Turku University of Applied Sciences
[3]Business, ICT and Life Sciences, Turku University of Applied Sciences
[4]RDI Services, Turku University of Applied Sciences
Joukahaisenkatu 3A, 20520 Turku, Finland

ABSTRACT

This paper presents the results from the field tests of two custom made exergames, based on a commercial off-the-shelf technology, aimed for the elderly people. First game resembles the guided mobility and stretching class, and the second one is more game-like with youthful theme and active movement. Both exergames were tested on two sites, urban and rural setting, totaling 19 elderlies. Usability findings are reported in a form of System Usability Scale (SUS) score analysis, and playability aspects as a Game Experience Questionnaire (GEQ) analysis. Further analysis is done from observation and interview material. The second game, which had more familiar setting and appropriate pacing, received positive feedback and higher scores from the tests. Based on this material we discuss about the design of exergames aimed for the elderly persons who are not experienced computer game players, the importance of graphical clarity and the need for specialized game experience questionnaire for the elderlies.

KEYWORDS

Exergames, Motion Tracking, Elderly, Serious Game.

1. INTRODUCTION

The number and proportion of the elderly in the population is growing, especially in the industrialized countries, and are one of the fastest growing age groups (UN, 2014). Age-related decline of physical and cognitive capabilities have severe impact on individuals quality of life as well as to the need for additional social and health related services. This combination leads to a rising costs on individual and societal levels. It is known that exercising the muscles (e.g. Orsega-Smith et al., 2012) and brain (Boot et al., 2013) helps to slow down, reverse and even stop the age-related decline (e.g. Anguera et al., 2013; Kuhn et al., 2014).

These physical and mental exercises can be done either independently or with an assistance (e.g. with physiotherapist or nurse). Common problems with either choice is the relapses from the chosen program due to repetitive and sometimes even boring nature of these exercises. People need ongoing motivation to keep on doing the repetitive tasks, and for this task the games and gamified solutions have been seen as a possible motivator.

In this article, two different exergames, a sports game and a gamified exercise game, are analyzed. Both games target to increase physical wellbeing using motion tracking techniques. The games were developed during a research project studying elderly gaming.

The main objectives of this paper are 1) to report the usability findings from SUS questionnaire, 2) to report the gaming experience felt by the elderly based on the GEQ questionnaire, interviews and observations, 3) to discuss what kind of features the elderly liked and disliked in these games and 4) to summarize the findings related to design decisions of these games.

The paper is organized as follows. Section 2 describes the games that were implemented for this work and the testing procedures used to obtain the results that are reported on Section 3. In Section 4 we discuss about the finding and the potential for the future studies.

ISBN: 978-989-8533-38-8 © 2015

2. GAMES AND RESEARCH METHOD

In this section we shortly describe how the exergames were designed and created, demographics of our participants and how the data collection was organized. Also, we will shortly introduce the games in question so that the readers have a better understanding on their similarities and differences.

2.1 Game Design and Creation

The games presented in this article have been developed as a part of the Gamified Solutions in Healthcare research project. Development of the games started already before the research project in 2013. The original goal was to create exergames that are played outdoors in an interactive public playground mainly by kids and youth. During the research project, the games were modified to be more suitable for elderly people. Players view to the games are presented in the following figure 1.

Figure 1. Game A (left) and Game B (right) from the player's point of view.

Game A. Originally this game was designed to activate big muscle groups while also being fun. In the game the player controls a boy with a scooter. The aim is to collect as many raspberries as possible while avoiding obstacles. The player can control character by moving himself in front of the camera. While the player moves left or right, character does the same. Character will move forward automatically and can also perform jumps and crouches if the player makes these movements. During the short level the speed of the character will gradually increase if he does not hit any obstacles. Hitting anything will also make the player lose some raspberries. After crossing the finish line, the player is given a number how many raspberries he managed to gather. For these test scenarios, mechanic that made character to lose speed and raspberries was disabled.

Game B. In the first concepts plans this game was a fast pace dance game. Development team changed the theme of the game for a current physical theme, because tracking dance moves with the selected technology did not work well enough. The aim of the game is to practice a set of physiotherapeutic gymnastic moves, engaged in a wide range of player movement. The moves are designed in collaboration with a physiotherapist while taking the constraints caused by the selected motion detection technology into account. A tempo and an atmosphere of the game are deliberately designed calmer than in the game A. This game works as a body maintenance exercise. The game gives most of the feedback for the player after the performance, so the player can focus on the exercise and the correct moves. A graphical presentation was to be considered similar to other games in the playground. This enabled the graphic designer to reuse previously made designs and graphics.

As mentioned, both of the games were modified to be more suitable for the elderly people. Game A was slowed down because original pace was considered to be too hectic for the target group. Because of previous testing, a function that made player lose a point when hitting on something was removed, so no- one would be left with zero points at the end of game. Also the jumping was tweaked to be so sensitive that the players could perform it by only nudging upwards. This allowed the users with balance problems to complete the game session without actually jumping upwards, which in some cases could be dangerous to their health. In Game B the user interface graphics needed a modification, as the original version had been a bit unclear. Otherwise the game was kept identical to the original.

2.2 Participants

Inclusion criteria for the study participant required that they are over 60 years old and are able to complete the games safely. This limited the recruitment of elderlies with severe physical disabilities or problems with balance. Also elderlies with advanced memory related problems were excluded as they might have problems on remembering what they were doing or were doing before the questionnaire part of the research protocol.

The testing was carried out in two different settings; rural and urban. First testing session was arranged in urban area with c. 200 000 inhabitants in South-Western Finland where the participants were recruited from an activity centers and associations for elderly citizens. Second testing session was arranged in rural area with c. 2500 inhabitants in Eastern Finland where the participants were recruited by contacting local pensioner's society, by advertisement in local newspaper and by asking from suitable persons who happened to be near the testing facility which was provided by a local service home.

2.3 Data Collection

To get quantitative data about the games and their features, we used two questionnaires; System Usability Scale (SUS) and Game Experience Questionnaire (GEQ). SUS is a short 10 item questionnaire used to collect usability where questions are rated on 5-point Likert Scale (Brooke, 1996). As we are examining exergaming, we modified the SUS questions to reflect this. Mainly, the word "system" was replaced with "game" and "use" by "play" with required conjugations. Additionally, to elucidate the SUS scores they were converted to grades by using the scale defined by Bangor et al. (2009).

GEQ (Poels et al., 2007) is a multipart questionnaire aimed on exploring the players emotions related to various aspects of the game in question. These aspects include their feelings of competence as players, was the experience negative or positive and how challenging the experience was. In total GEQ has 50 questions for players that played the game alone, and 67 to players who played game with one or more partners i.e. in a multiplayer setting. Main parts of the questionnaire are the Core Module, Post-Game Module and the Social Presence Module for multiplayer settings. Components in these modules have a range from 0 to 4, and they represent the players' self-reported view of how they felt or experienced various aspects during the gameplay. These experiences can be positive ("I felt skillful") or negative ("I felt frustrated").

In addition to the quantitative data, all the sessions were filmed from two angles and observed by several members of the research team. First camera was situated behind the player so that it could capture the players' movements and the action on the screen simultaneously. Second camera recorded the movements and expressions of the player from the frontal view. This material was used to observe how well the players could follow the tasks presented on the games. The recordings were also used to inspect how the players used their bodies in different situations and how confident they were on their movements and actions. Following figure 2 depicts the settings in Urban setting on the left, and in the Rural setting on the right.

Figure 2. Setup of the gaming sessions, Urban (left) and Rural (right).

All the sessions were observed by participating research team. In the Urban setting two of the team members were guiding the gaming part of the session and helped the participants with the filling of the questionnaires, while one member was observing the gaming part and doing the interview after that. In the Rural setting two of the team members were guiding the gaming part of the session, one helped the participants with the filling of the questionnaires, and two additional members were observing the gaming session from the side. These qualitative observations are used to support the interpretation of the questionnaire results.

After the gaming session was over, the players answered to series of yes/no questions that acted as a themes for short interview. These questions were based on the quantitative STAM questionnaire (by Renaud et al., 2008) as it gave us a well-reasoned background related to our focus group. STAM itself was not used as it was decided that two long questionnaires would take too much time to fill for the elderlies and also the time used to fill the questionnaires versus time the games were played was already heavily geared towards the data collection. With these questions we probed the elderlies to get their own opinions and thoughts about the games they just had played.

3. RESULTS

In total we had 19 participants with overall average age 72±11 years (n=19). In the urban setting we had seven participants (n=7, 3 female and 4 male), and in the second testing session in the rural settings we had 12 (n=12, 7 female and 5 male) participants. Age distribution was similar in both settings, and both genders were represented equally (10 female, 9 male).

All the participants were active elderlies who live independently, and had physically active life. In average they had light or moderate exercise for 1½ hours per day, including some guided exercise groups or activities. 3 out of 19 participants reported on playing Solitaire occasionally, and one of the participants used to play puzzle games on tablet computer. Otherwise digital gaming as an activity was something they had not tried ever, or they had just seen how their grandchild's were playing something.

Reported health related problems in this group included variety of problems with knees, hips and hands. Most severe problems were heart attacks which mean that these participants could not raise their hands too high for prolonged periods of time or at all, breast cancer operation, and a shoulder operation which prevented the use of the hand in question as the operation was quite recent. Other mentioned problems included mild depression and cases of epilepsy. Participants did not report memory related problems or diagnoses.

3.1 System Usability Scale Scores

Previously a sample of at least 12 participants has been calculated as minimum to obtain sufficient data on the SUS questionnaire (Lewis et al., 2009). Our sample had an n = 19, so the sample size for the combined results fill this requirement for the reliability. The mean SUS score for Game A was 58.29 Ok/F, and for Game B 79.44 Good/C. Following table 1 shows the SUS scores and grades for both games.

Table 1. SUS scores

	SUS	Min.	Max.	Median	Grade
Game A	58.29	30	80	62.5	Ok / F
Game B	79.44	45	100	81.25	Good / C

Traditionally product evaluated with SUS should achieve around 70 points from the test to be considered to be ‘passing the test’ (e.g. 70 points in Bangor et al. 2009; 68 points in DoHHS, 2014). When we use these limits to evaluate Game A and Game B, Game A fails to reach this level of minimum acceptable usability with this group of users.

When we analyze the SUS score on the level of individual questions, we find some key differences between our games. Game B gains its advantage on three questions; 3. “I thought that playing the game was easy”, 8. “I found the game very cumbersome to use” and 9. “I felt very confident while playing the game”. In these questions Game B was seen in much more positive light, while Game A was seen as cumbersome and complex.

3.2 Game Experience Questionnaire

GEQ has three parts that probe on different aspect related to the players experience with the game. The Core module which is split on seven components represents the players' self-reported experience during the gaming session. Results of this module are presented on table 2.

Table 2. GEQ Core

	Competence	Sensory and Imaginative Immersion	Flow	Tension / Annoyance	Challenge	Negative affect	Positive affect
Game A	1,47	2	1,85	0,49	1,29	0,3	2,4
Game B	2,3	2,36	2,25	0,14	0,78	0,07	2,75

Game B has favorable scores on all the seven components. It gets higher scores on four components that have positive connotations attached to them, and lower scores on the negative ones. Especially Game B receives high score on the Competence factor which represents the players feeling about how well she was doing during the gameplay. Challenge is a component that can be either positive or negative. In the case of Game B the score is lower, which can be interpreted that game was easier for the participants and/or that thanks to other factors like Competence, the participants did not notice the Challenge so much as with Game A. Overall, Game A was deemed to be more frustrating than Game B.

The Post-game module measures how the players feel about the game after the gaming session has ended. This module has four components on it, and these are presented on table 3.

Table 3. GEQ – Post-game module

	Positive experience	Negative experience	Tiredness	Returning to Reality
Game A	1,76	0,23	0,31	0,6
Game B	1,989	0,06	0,21	0,5

Returning to Reality component represents the players return from the immersed state to the out of the game world. In this case the gaming sessions were short (typically under 15 minutes) and game world was not designed to be the immersive and total experience. These factors might render this component to be irrelevant, but Game A received higher value indicating that participants were more immersed on it. Game B is again coming out as more positive experience than Game A.

Social Presence module is used when game in question is played with co-player or players. This module measures how the other player(s) affected the behavior and feelings of the questioned player. These results are presented on the following table 4.

Table 4. GEQ – Social Presence module

	Psychological Involvement – Empathy	Psychological Involvement – Negative Feelings	Behavioral Involvement
Game A	1,5	0,32	0,45
Game B	1,323	0,51	0,3

Regarding the Social Presence both games scored quite low scores, but Game A had slightly better social impact. Accordingly, both factors related on the feelings towards the other player and their actions during the gaming session generated low scores. The observations during the gameplay confirmed that participants did not give notes to each other or learnt what to do to improve their performance in the game from each other. Observations also revealed that these games did not create shared gaming experience for the participants.

This might be due the fact that the gaming sessions were short, so the participants did not have enough time to start commenting the actions of the other player or to take notes on what they were doing.

3.3 Interviews

The interviews done after the gaming sessions were based on a questions influenced by the STAM framework (Chen, 2014) and were modified to better fit the domain of exergames and elderlies. The interview was conducted by first asking a simple yes/no question related to the games that were tested, and then elaborating a short discussion the question topic. The results of the yes/no questions are on the following table 5.

Table 5. Qualitative results

		Yes	No
Perceived usefulness	Could the playing of games be an efficient and easy way to do exercises?	17	3
Perceived ease-of-use	Was playing the games easy?	15	5
Gerontechnology self-efficacy	Could you play the games after you got the instructions? Would the instructions manual be enough?	17	4
Gerontechnology anxiety	Were you scared of making mistakes while you were playing the games?		19

The total amount of answers per question is higher than the number of participants we had. This is due the fact that many of them gave two answers, one for both games as they had differing experiences from them. In the case of perceived usefulness, the elderly thought that games could be useful way to do exercises. But, they almost unanimously preferred to do their exercises either on gym-like environment or outside their apartment. Exception to this was that several of them said that games could be useful when the weather conditions are bad and therefore it is not possible to exercise outside.

Despite the reported negative feelings and frustrations related to the gaming sessions or notes about how they do not know how to use computers, after the gaming session our participants did not think that these games were difficult for them. In general the perceived ease of use was favorable, but some elderlies said that games are not designed for elderlies or for especially for them as individuals.

In the theme falling under the gerontechnology self-efficacy, most of our participants said that they could use the games without help, but it would create "of course" more problems and they would have to learn new skills. But they did not see this as a major problem and thought that they could learn to use the system after the initial learning period. Some of the elderly thought that they would require an assistance on the usage of this kind of a system at home as they were not accustomed to use technology.

In spite of the challenges with technology, games and health, gerontechnology anxiety was unanimously dismissed. Our participants did not fear the technology or about making mistakes with the games, instead they noted that the 'fear' they felt was about losing points in the game.

The participants were also given a chance to tell their own opinions freely about the gaming session they had participated on. On these comments they commented on how in general exercising is a good thing for their age group as it keeps them active and also helps them to stay healthy. The gaming as an activity was commented to be something that they had not tried and experienced previously, but it had been an interesting experience nevertheless.

4. CONCLUSION

Our goal is to activate elderlies that are not already actively engaged in physically demanding activities. To achieve this goal, we have built on the experiences of previous projects done by others (e.g. such as Join-in, 2011) and by us. Game A and B have been our first games that were tested with the participants that belong to our targeted age group. From these games we have learned what kind of things seem to work with our target group and what does not work. To remedy the problems we found during the testing, we are currently doing a redesign on both games.

We had designed the games to have clear graphics without a screen clutter which is seen on games aimed for younger players. Despite our efforts, the participants still had problems with knowing what they should avoid and what they could collect on the Game A. The graphics problem might be partially caused by the unfamiliarity of the activity represented in the game A as the interviews and comments by the elderlies during the gameplay revealed their confusion about these matters. The speed in which things happen on the Game A was also a problem, albeit the game was significantly slower than the original version which was designed for the younger players. The next version of the game will have its pacing further slowed down, and the context of the game will be more familiar to the target group.

Game B had originally a virtual instructor whose movements' player should match, and a graph that showed what kind of movements will be coming after the current one. This caused confusion as some of the participants were not sure which one to follow, or they missed the correct movements because the extra graph grabbed their attention. In Game B players have to raise their hands several times above their heads. This was a problematic procedure for several elderlies, and prevented them from completing the required movements. These problems are addressed in the new version of the game where e.g. the indicator for the upcoming motion has been removed.

In the future field tests, the gaming sessions will be longer as it was noted that during the short sessions the participants were aware about their surroundings and the observers. The longer sessions might help the participants to relax and immerse more on the game, instead of them being filling questionnaires in every turn after few minutes of the play. This might allow us to see more emotions rising from the gameplay experience.

SUS is a simple tool for measuring basic usability of a software-based artefact, and it proved to be usable and informative also with the elderlies, mainly thanks to its brevity. GEQ in other hand was a questionnaire which contains 50 or in case of multiplayer situation 67 questions. The length of the GEQ brought up questions and comments from the participants during the testing sessions (e.g. "When this questionnaire will end..?", "How many questions there are left?"). In future the game experience research with elderlies, or with other people who do not play games regularly or at all, would benefit from a shorter and more understandable questionnaire.

ACKNOWLEDGEMENT

This article was done as a part of the Gamified Solutions in Healthcare research project. The project is conducted by University of Turku and Turku University of Applied Sciences together with partners Puuha Group, Serious Games Finland, City of Turku and Attendo. The project is funded by Tekes – the Finnish Funding Agency for Innovation. Both games in this article are using the Extreme Motion technology by Extreme Reality Ltd. (http://www.xtr3d.com) and Unity game engine by Unity Technologies (http://unity3d.com).

REFERENCES

Anguera, J.A.J., Boccanfuso, J. L., Rintoul, O., Al-Hashimi, F., Faraji, J., Janowich, E., Kong, Y., Larraburo, C., Rolle, E., Johnston and Gazzaley, A., 2013. Video game training enhances cognitive control in older adult. *Nature*, 2013.

Bangor, A., Kortum, P., & Miller, J., 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, Vol. 4, pp. 114-123.

Boot, W.R., Champion, M., Blakely, D.P., Wright, T., Souders D.J., and Charness, N., 2013. Video games as a means to reduce age-related cognitive decline attitudes, compliance, and effectiveness. *Frontiers in Psychology*, 2013.

Brooke, J., 1996. SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. *Usability evaluation in industry*. London: Taylor and Francis; 1996. pp 189–194

Chen, K. & Hoi Shou Chan, A., 2014. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM), *Ergonomics*, Vol. 57, Issue 5, pp. 635-652.

Join-In project, 2011. Join-In Senior Citizens Overcoming Barriers by Joining Fun Activities. Project deliverables of Join-In EU FP7 -project, 2011. Material also available on: http://joininproject.wordpress.com/

Kuhn, S., Gleich, T., Lorenz, RC. Lindenberger, U., and Gallinat, J., 2014. Playing Super Mario induces structural brain plasticity gray matter changes resulting from training with a commercial video game. *Journal of Molecular Psychiatry*, 2014.

Lewis, J., Sauro, J., 2009. The factor structure of the System Usability Scale. In: Kurosu M, editor. *Human centered design*. Heidelberg, Berlin: Springer; 2009. pp 94–103.

Orsega-Smith, E., Davis, J., Kelley Slavish, K. and Laura Gimbutas, L., 2012. Wii Fit Balance Intervention in Community-Dwelling Older Adults. *Games for Health Journal*, 2012.

Poels, K., deKort, Y. A. W., & Ijsselsteijn, W. A., 2007. Game experience questionnaire. *Project deliverable for the EU IST project the Fun of Gaming*.

Renaud, K., van Biljon, J., 2008. Predicting technology acceptance and adoption by the elderly: a qualitative study. *Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology*, ACM, Wilderness, South Africa (2008)

United Nations, Department of International Economic and Social Affairs, Population Division, 2014. *Concise report on the world population situation 2014*. ST/ESA/SER.A/354. ISBN 978-92-1-151518-3. New York, 2014.

U.S. Department of Health & Human Services, 2014. Website: http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

TOWARDS AN UPDATED TYPOLOGY OF NON-PLAYER CHARACTER ROLES

Henrik Warpefelt and Harko Verhagen
Department of Computer and Systems Sciences, Stockholm University
Borgarfjordsgatan 12, SE-16407 Kista, SWEDEN

ABSTRACT

In video games, non-player characters (NPCs) provide important services in that they facilitate the player's interaction with the game in a way that is in accordance with the expectations set by the narrative. It is, however, still unclear in what ways these NPCs must act, look, and feel in order to fulfill these expectations. In this study we aim to establish a typology of the roles NPCs play in games, building on a previous typology by Bartle (2004) aimed at providing a framework for describing the requirements put on NPCs by these expectations. This was done via an online survey, where respondents were asked to classify NPCs in images from 4 games, and to provide a description of why they classified it as belonging to a certain role. The results of the survey were the analyzed for instances where players expressed confusion about which role an NPC belonged to. These findings were used to update the previous typology. The results from this were later verified by applying the new typology to 10 other games. In the end we identified a number of new roles, as well as modifications to existing roles, which when combined with Bartle's original typology created a typology applicable to a larger number of genres.

KEYWORDS

Non-player characters, roles, affordance, interaction, games, evaluation

1. INTRODUCTION

Within video games, non-player characters (NPCs) provide many functions to players (Bartle, 2004). They not only provide challenges, but also allow the player to manage resources, provide narrative exposition, and make the world feel alive. In many of these roles, NPCs provide direct game functions to the player, for example buying or selling goods (Bartle, 2004). In order for the player to be able to utilize these functions, they first need to be able to perceive them (McGrenere & Ho, 2000).

In a previous study (Warpefelt, 2015) we performed a survey where respondents were asked to classify examples of NPCs from 4 games as belonging to different roles, and to provide the reasoning for their decision. This study showed that the roles established by Bartle (2004) were somewhat lacking in descriptive power for some types of games. In this study we aim to remedy this, and to that end we updated Bartle's list of roles using the results gathered from the previous study (Warpefelt, 2015) and verified this by applying the new typology to 10 other games from diverse genres. In the end, we identified a number of new roles, as well as a number of modifications to Bartle's list.

2. BACKGROUND

NPCs play a critical role in upholding the believability of a game world, both by their behavior but also by their appearance (Loyall, 1997). By making the game believable, they also strengthen the player's feeling of immersion (Johansson 2013; Johansson, Warpefelt & Verhagen, 2014). However, the visual representation of NPCs is not only defined by the functions they provide within the game, but also encompasses and makes perceptible the roles they play in the narrative of the game. Therefore, the visual representation of an NPC must not only signal its functional role, but also be in line with the expectations set by the narrative (Loyall, 1997; Desurvire, Caplan and Toth, 2004). NPCs take on many different forms in games. In some cases, they take on human-like forms and interact with the player in peaceful ways. In others, they are monsters who exist to fight the player (Bartle, 2005). In less common cases, they may take on forms not commonly associated with characters, for example vehicles or trees. The distinction used in this work is if they achieve some level of *characterhood* – i.e. the active portrayal of a character as opposed to merely existing in the world (Warpefelt, 2013). For example, a shopkeeper would qualify as an NPC, whereas a vending machine would not. This distinction can sometimes be difficult to make, and the level of behavior needed to achieve characterhood could come down to very slight differences. In the shopkeeper example above, giving the shopkeeper a spoken voice and making them look like a shopkeeper would push them over the edge, whereas the vending machine would have to perform actions that would be truly extraordinary for a vending machine (for example be able to hold a conversation) to qualify for characterhood. In order for the NPC to achieve characterhood, and for the game to maintain believability and enjoyment, however, the representation of the NPC must be presented in accordance with the roles that the player has understood the NPCs to have in the game (Loyall, 1997) as they are signaled by the narrative and the created expectations. In many cases, the achievement of characterhood will come down to the player's expectations on the character, and to what extent the player's imagination will "fill in the blanks". For example, a bus may qualify as an NPC in a racing game if it were actively hindering the player from completing the race since the player would likely assume that a moving bus would have someone at the wheel. The bus thus becomes the avatar of the virtual bus driver, as it were.

2.1 Types of NPCs

Over the last decade, a number of ways of describing the roles provided by NPCs have been presented. In many cases, these are often focused on in what ways NPCs fit into the narrative of the game, such as the work of Aarseth (2012). Aarseth's typology encompasses many aspects of the game narrative, and NPCs are categorized depending on the depth of their character.

Lankoski and Björk (2007, 2008) have famously introduced a number of design patterns and principles for NPCs. However, these are aimed at how one can design believable characters and although this is valuable information, it does not constitute a typology as such.

Bartle (2004) provides typology of NPCs which lists a number of roles that NPCs take on in games, where each role represents a certain function provided by NPCs within the game. These roles are derived from Bartle's research into multi-user dungeons (MUDs) and are by-product of his design work. They are:

- Buy, sell and make stuff
- Provide services
- Guard places
- Get killed for loot
- Dispense quests (or clues of other NPCs' quests)
- Supply background information (history, lore, cultural attitudes)
- Do stuff for players
- Make the place look busy

Each of these roles defines a function provided by an NPC within the game, as opposed to in what ways the NPC ties into or supports the narrative. However, the NPCs still exist within the narrative and therefore this role must also be represented in such a way that the player is able to perceive and interpret the NPC as being of a specific role.

2.2 NPC Affordances and the Narrative

In order for the functional roles to be perceptible to the player, the NPCs that provide them must tie into the narrative and present indicators as to which roles the NPCs pertains. These indicators are usually small cues that tell the player what to expect of the NPCs (Fernández-Vara, 2011). This process not only feeds into the affordances (Gibson, 1977) of the NPC, but also as previously mentioned the creation of the narrative. As the player plays the game, they encounter numerous areas, items, and NPCs and each of these feed into what Calleja (2009) calls the *alterbiography*. Calleja's alterbiography is essentially the narrative that arises as the player experiences the game. This is contrasted with the *scripted narrative*, which consists of pre-made sequences providing exposition of the overarching story of the game. The scripted narrative becomes a part of the alterbiography, albeit in the way the player *experienced* the story rather than how the story was presented. However, the scripted narrative also provides the framework through which the player understands the world, and thus creates their alterbiography.

Table 1. Games included in the initial study, sorted by title

Title	**Developer**	**Year**	**Description**
Assassin's Creed: Unity	Ubisoft	2014	Alternative history sandbox action adventure game
inFamous: Second Son	Sucker Punch Studios	2014	Science-fiction action adventure game
WATCH_DOGS	Ubisoft Montreal	2014	Science-fiction sandbox action adventure game
World of Warcraft: Warlords of Draenor	Blizzard Entertainment	2014	Fantasy massively multiplayer online role playing game

3. PLAYER PERCEPTIONS OF NPC ROLES

In a previous study (Warpefelt, 2015) we performed an online survey to study in what ways respondents identify the functional roles of NPCs. In the initial study, some respondents also indicated that the Bartle roles were insufficient to describe the roles of the NPCs.

3.1 Survey Design and Respondent Demographics

The survey consisted of two parts. The first part collected basic demographic data about the respondents. In the second part, the respondents were presented with images of NPCs in games, and asked to determine which Bartle role they belonged to, or to suggest another role to which the NPC in the image belonged. The respondents were also asked to provide a description of their reasoning, and if they were previously familiar with that particular NPC. The images were captured in 4 games, as seen in Table 1. Each respondent was presented with 10 images randomly selected from a pool of 27 images. The previous study was aimed at identifying the criteria by which players evaluate NPCs using visual characteristics. In the study, we found that players identify the role based on the surroundings area and locations of the NPC, actions taken by the NPC and the attributes and visual appearance of the NPC. The survey had a total of 294 responses, of which 213 were complete and the respondents where aged 18 or older. For the valid responses, respondent age ranged from 18-52 years, with a mean of 26.3 ± 6.5 and a median of 21.5. Respondents were primarily from Sweden (29%) and the United States (29%), with the remainder mostly hailing from various western countries. The valid responses had a gender distribution of 76% male, 21% female and 1.5% other. 1.5% declined to specify their gender.

3.2 Survey Analysis and Results

For this study, the results of the survey were analyzed using thematic analysis in order to identify situations where players were unsure of which of the NPC roles laid down by Bartle (2004) may be applicable to the particular NPC. The themes identified were respondents explicitly being unsure of what role the NPC belongs to, actively discussing the roles in their response, and expressing that more than one may be suitable.

In performing the study, we discovered that although Bartle's typology has some descriptive power, it may be over-adapted to the MUD games it was originally based on. As a result of this, some of the Bartle roles need to be revised to better fit modern games.

3.2.1 Buy, Sell and Make Stuff, Dispense Quests, and Provide Services

The respondents to the survey often found it difficult to differentiate NPCs that provide services, buy/sell goods, and quest givers. In many cases, the respondents would mention that they needed user interface (UI) components would differentiate these roles of NPCs. Nevertheless, these roles are distinct and should be regarded as belonging to different roles since they let the user access different functions within the world. In the previous study, NPCs of these roles were identified primarily by their locations and actions – NPCs found standing passively in shops and easily accessible areas were generally assumed to be vendors, quest givers or service providers. In places where there were items seemingly for sale, NPCs were primarily assumed to be service providers or quest givers. NPCs were primarily identified as quest givers if they somehow stood out from their surroundings – for example if they were dressed differently.

3.2.2 Get Killed for Loot and Guard Places

The respondents also found the *Get killed for loot* role confusing, and pointed out that not all NPCs that are killed actually provide loot. In addition, many respondents noted that NPCs who *Guard places* often fulfilled other roles as well. NPCs of these roles were invariably identified as having generic or non-unique appearances – they were essentially generic entities stamped from the same mold.

3.2.3 Do Stuff for the Players

There seemed to be a general confusion as to what functions an NPC of this role would provide. Some respondents interpreted this as being NPCs that will help a player fight battles, whereas others would interpret this role as being NPCs that will guide the player to areas of interest or help them carry items. NPCs of this role were primarily identified by their behavior, in that they were actively assisting the player.

3.2.4 Supply Background Information and Make the Place Look Busy

The roles of *Supply background information* and *Make the place look busy* were perceived as similar by the respondents. Functionally, they both provide background information in that they influence the creation of the alterbiography, but the act of passively supplying background information was sometimes confused with making the place look busy.

3.2.5 Overall comments

Many respondents seemed to find it difficult to decide on the roles of some NPCs since they could potentially belong to more than one role. For example, vendors in some games would also provide repair services. This leads us to believe that NPC roles should not be viewed as either-or selections, as in the previous study, but rather as roles of which the NPCs can assume one or more.

3.3 Revised NPC Roles

Based on the data gathered from the survey, there seems to be some level of mismatch between in what ways the respondents interpret NPC roles and which are available in Bartle's typology. This leads us to believe that there is a need for an updated typology.

Although killing NPCs for loot is common in many role playing games, this type of resource gathering is not present in many other genres of games - where the NPCs instead only provide a challenge to the player. Because of this we propose that the *Get killed for loot* is split into two roles: *Provide combat challenges* and *Provide loot*. The prior role will also encompass enemies that aren't necessarily killed outright, for example enemies that surrender or are defeated using non-lethal methods.

Furthermore, the *Do stuff for players* role is very vague, and should ideally be split in two, where the NPCs that help the player in combat fall under the *Help the player* role, whereas NPCs who provide the player with non-combat assistant fall under *Provide services*. With these changes in mind we propose the following typology:

- Buy, sell and make stuff
- Provide services
- Guard places
- Provide combat challenges

- Provide loot
- Dispense quests (or clues of other NPCs' quests)
- Supply background information (history, lore, cultural attitudes)
- Assist the player
- Make the place look busy

4. VERIFYING THE NPC ROLES

The NPC roles identified above were then verified in an extended study, aimed at reaching a more diverse basis for the data material. The games included in this part of the study can be found in Table 2. The games were selected for diversity among genres, with focus on game play genres rather than narrative genres.

4.1 Data Collection

The functions provided by NPCs in the games were elicited by analyzing game play from the games using coding, where NPCs were coded as belonging to the roles defined in the previous step using the criteria from the previous study. Game sessions were captured as video. This was performed by one researcher. Each game was played until theoretical saturation was reached, but no less than a total of two hours. This included a familiarization period, where the researcher explored the game to acquire an understanding of the game. These familiarization sessions were usually not recorded, but counted towards the total time in-game. If the researcher was already familiar with the game, the familiarization period was sometimes skipped. The familiarization period was then followed by an actual play period, where researcher played the game normally and recorded the session. In doing the analysis, only NPCs who were perceived as having achieved characterhood were included in the data collection.

4.2 Data Analysis

The game sessions were later analyzed using thematic analysis to identify the functions provided by the NPCs found within the games. For each game we identified functions provided by NPCs, along with the characteristics of the ways in which the NPCs provided these functions. The resulting functions from each game were then mapped into the previously established roles. In cases where the provided function could not be mapped to a distinct role, the function was noted and saved for later cross-game analysis.

Each of the intermediate roles were found at least twice across all the games, but only in two cases *Skyrim* and *The Old Republic*) were all the roles found in a single game. Below we will discuss each role in turn, as well as any functions that could not be mapped to a distinct role. During this analysis, only situations where the player could control the avatar were considered.

Table 2. Games included in the verification study, sorted by title

Title	Developer	Year	Description
Bioshock Infinite	Irrational Games	2013	Science-fiction first person shooter
Burnout Paradise	Criterion Games	2008	Modern-day racing game
Company of Heroes	Relic Entertainment	2006	Historical real-time strategy game
Dungeon Defenders	Trendy Entertainment	2011	Fantasy tower defense strategy game
L.A. Noire	Team Bondi/Rock Star Leeds	2011	Modern-day murder mystery game
Saints Row: The Third	Volition	2011	Science-fiction open-world action game
Star Wars: The Old Republic	BioWare	2011	Science-fiction massively multiplayer role-playing game
Tomb Raider	Crystal Dynamics	2013	Modern-day action-adventure game
The Elder Scrolls V: Skyrim	Bethesda Softworks	2011	Fantasy role-playing game
Tropico 3	Haemimont Games	2009	Historical/modern-day strategy/building game

4.2.1 Buy, Sell and Make Stuff and Provide Services

NPCs filling these roles showed up in only three games in the study, namely *Dungeon Defenders*, *Skyrim*, and *The Old Republic*. In all three games, many of the NPCs which filled the role of *Buy, sell and make stuff* also fulfilled the role of *Provide services* in that they would often give the player the ability to repair or upgrade items, or in the case of *Dungeon Defenders* sell let the player buy bonuses that allowed them to advance faster or change previous in-game choices.

4.2.2 Guard Places

NPCs acting as guards appeared in five games: *Bioshock Infinite*, *Tomb Raider*, *Skyrim*, *Saint's Row*, and *The Old Republic*. Across the game, guarding behavior comes in two forms: either they're guarding a location or person with the express purpose of denying the player access to them, or they are guarding the player or their territories against the player's enemies. In the prior case, these NPCs invariably filled the role of *Provide combat challenges*, whereas in the latter case they filled the role of *Assist the player*. This distinction makes the role of *Guard places* seem superfluous, and it could easily be subsumed into the other roles. This difference is largely behaviorally based, and since both types can appear in similar contexts it would be pointless to try to differentiate the two roles based on visual attributes or environment.

4.2.3 Provide Combat Challenges and Provide Loot

NPCs that provided combat challenges were found in all games except for *Burnout*. In all cases, these NPCs perform the role in essentially the same way: by attacking the player or their representatives with mêlée or ranged weapons. In many cases, the NPCs fulfilling this role would also *Provide loot*, which was evident in all games except for *Company of Heroes*, *Tropico 3*, and of course *Burnout*. Although the frequency with which these two roles are paired lend credence to Bartle's grouping of them, there are still examples that warrant their separation into separate roles - in this case the strategy games *Company of Heroes* and *Tropico 3*. NPCs providing combat challenges will, as with the previous type *Get killed for loot*, largely be faceless copies. NPCs providing loot are slightly different and will often provide items they were holding or using in first person shooters and action games, but can provide unexpected items in role playing games – for example animals providing money.

4.2.4 Dispense Quests (Or Clues of Other NPCS Quests)

This role was primarily found by NPCs in *Skyrim* and *The Old Republic*. A case could be made for their appearance in *LA Noire* and *Saint's Row*, but in these cases the NPCs are mostly present when the assignment is given out rather than them dispensing it in person. In *LA Noire*, the player's commanding officer will give the player mission descriptions in the form of in-game cutscenes, but never when the player actually has control of their avatar.

Furthermore, this role is somewhat awkwardly named and should ideally be shortened to *Give or advance quests*. This also fits better with modern games, where quests often take the form of chains rather than singular objectives. The evaluation criteria should remain the same.

4.2.5 Supply Background Information (History, Lore, Cultural Attitudes)

Supply background information presents an interesting problem, since it does not differentiate what Calleja would call the predefined narrative and the alterbiography (Calleja, 2009). If one would interpret this as being related to the predefined narrative, it appears in all games except for *Dungeon Defenders*, *Company of Heroes*, and *Burnout*. If one were to interpret it as meaning the building of the alterbiography, it essentially shows up in *all* games where there are NPCs since they will invariably help build the alterbiography. In the second case the role becomes meaningless as it will be applicable to all NPCs and thus lose all analytical power. Essentially, this role should only encompass NPCs that actively provide background information.

Essentially, this role needs to be more clearly defined and renamed to reflect its actual purpose. We suggest that the role be delimited to only cover the predefined narrative, and that it be renamed *Provide narrative exposition*. The evaluation criteria should remain the same.

4.2.6 Assist the Player

This role appeared in *Bioshock Infinite*, *Tomb Raider*, *Skyrim*, *Saints Row*, and *The Old Republic*. It essentially takes on two aspects: one where the NPCs will fight with the player, and one where the NPCs simply guide or advice the player. These are very different purposes, and they should be represented as different roles.

In combat situations, we can identify two separate roles: one where the NPCs act autonomously to help the player, for example police officers in *The Old Republic* helping the player fight enemies if they attack them close to where the police officer is. The second role is one where the NPC will follow the player around and assist them - essentially accompanying them as a follower. In this latter case, the NPC can apply a certain level of control of the NPC, for example telling them to be aggressive or passive. We suggest that these roles be called *Act as an ally in combat* and *Accompany the player*, respectively. *Act as an ally* would be NPCs that will autonomously fight with the player, and *Accompany the player* will be NPCs that will fight with the player under their control.

In non-combat situations, NPCs will often provide help, directions, advice, and resources to the player. This is done autonomously, and the player rarely has to react to receive this help. This role could feasibly keep the name *Assist the player*.

4.2.7 Make the Place Look Busy

This role appeared in all games except for *Dungeon Defenders* and *Company of Heroes*. In these two games, all NPCs present have a role to fulfill, and they are either on the player's side or on the other side. They exist for some express purpose, and not just to make the place look busy. In games where this role is found, this role is often represented by NPCs that have very limited interaction capabilities, although many special-purpose NPCs in the game will also act in such a way that the world seems more populated. As mentioned in section 4.2.5 (*Supply background information)*, all NPCs can be used to augment the alterbiography of the player. Instead, this role is intended to differentiate NPCs whose function is to act as an extra. They provide no other services except for making the world look busy.

4.2.8 Functions not Matched to Roles

In *LA Noire* and *Burnout*, we identified a number of situations where NPCs would provide challenges that were not related to combat. In *LA Noire*, this primarily took the form of interviews, whereas in *Burnout* it took the form of vehicle races. In both of these cases, the primary task of the player was to overcome a challenge set by the game mechanics rather than one of combat. Although this could be collapsed into the *Provide combat challenges* role, it is a fundamentally different role. Thus, we suggest that a *Provide mechanical challenges* role be added. This could potentially also be extended to other mechanical challenges in games, such as NPCs affecting the environment in movement puzzles. NPCs that provide mechanical challenges will likely be similar in appearance and placement to NPCs that *provide quests* or *buy, sell and make stuff*.

In *The Old Republic* we also encountered player pets, which is essentially a purely decorative NPC that follows the player around. It has little to no impact on game play, and is simply there as a vanity item. This could seemingly be folded into the *Accompany the player* role, but doing so would go against the purpose of that role in that it is supposed to define companions who help the player. Instead, we suggest that this type of NPC be encompassed by the *Make the place look busy* role.

4.3 Suggested NPC Functional Roles

Based on the findings outlined above, it is evident that the roles originally proposed by Bartle suffer from over-fitting to the role-playing genre and are incapable of capturing the functions provided by NPCs found in many other genres. Therefore, we recommend that the list of NPC roles be further modified to better encompass a broader spectrum of games. We propose that the following list of roles be constructed as follows:

- Buy, sell and make stuff
- Provide services
- Provide combat challenges

- Provide mechanical challenges
- Provide loot
- Give or advance quests
- Provide narrative exposition
- Assist the player
- Act as an ally in combat
- Accompany the player
- Make the place look busy

It should be noted than although these roles are different – they can be portrayed in unison. It is not unfeasible that an NPC would “give or advance quests” as well as “buy, sell or make stuff”. Similarly, most NPCs will make the place look busy simply by existing.

5. CONCLUSION AND DISCUSSION

In this paper we identified a number of new roles, as well as a number of modifications to Bartle’s existing roles, that should be used to form a new typology based on Bartle’s original list of roles. From this work it can be concluded that although Bartle’s roles are very well suited at describing role playing (online or otherwise) games, especially older ones, they suffered from a number of weaknesses in regards to generalizability both outside the role playing genre, but also to newer forms of games – most pressingly massively multiplayer online role-playing games.

The roles that we have identified also have a greater level of separation than those provided by Bartle, and should be more useful as analytical tools for both game researchers and game designers. By connecting actual functional roles and affordances to behavior and presentations, we can begin to understand and to some extent formalize the design language used by game developers and evaluators. Although this formalization carries with it the risk of homogenization of games and game content, it provides a concrete benefit in that it can further the dependability of user interfaces in games, and user interaction with games.

It should be noted that Bartle’s initial typology was derived from design work, whereas the typology presented above is derived from studying the perceptions of players. Although these are different ways of approaching the problem, both are essentially player-centric approaches. If nothing else, NPCs must be designed in ways that allow them to convey their function(s) to players.

5.1 Future Research

Although this study identified a list of NPC functional roles, these should be seen as a first step towards a full-scale typology. The current typology still needs to be further verified, ideally against a broad base of respondents and using a larger sample of games. Given the small sample for this study and the previous one, there exists a risk of over-fitting similar to the one evident in Bartle’s typology. There may also be cases where numerous roles arise in cohesion and are difficult to separate – as seen with some of the roles from Bartle’s typology. Through verification with different groups of respondents we could also identify and document potential differences in interpretation between respondent groups – something that may arise as an issue since games are a cross-cultural phenomenon.

Lastly, although the roles we describe have basic descriptions and identifying characteristics, there will be a need for the roles to be further described and key components of them identified – ideally providing a large complement of traits for each of the characteristics we identified in the original study (Warpefelt, 2015), namely the surroundings area and locations of the NPC, the actions taken by the NPC and the attributes and the visual appearance of the NPC. For this typology to become truly useful as a design and evaluation tool, a library of examples and typical expressions of each characteristic would be useful.

REFERENCES

Aarseth, E., 2012. A narrative theory of games. *Proceedings of the International Conference on the Foundations of Digital Games*. New York, NY, USA, pp. 129–133.

Bartle, R., 2004. *Designing virtual worlds*. New Riders Publishing, Indianapolis, Indiana, USA.

Desurvire, H., Caplan, M., and Toth, J. A., 2004. Using heuristics to evaluate the playability of games. *CHI'04 extended abstracts on Human factors in computing systems*, Vienna, Austria, pp. 1509–1512.

Ermi, L., and Mäyrä, F., 2005. Fundamental components of the gameplay experience: Analysing immersion. *Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in play*. Vancouver, British Columbia, Canada.

Fernández-Vara, C., 2011. Game spaces speak volumes: Indexical storytelling. *DiGRA 2011 - Proceedings of the 2011 DiGRA International Conference: Think Design Play*. Utrecht, the Netherlands.

Gibson, J. J., 1977. The Theory of Affordances." *R. E. Shaw & J. Bransford (eds.), Perceiving, Acting, and Knowing*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

Gordon, C., 2009. Experiential narrative in game environments. *DiGRA 2009 - Proceedings of the 2009 DiGRA International Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory*. West London, UK.

Johansson, M., 2013. *Do Non Player Characters dream of electric sheep?: A thesis about Players, NPCs, Immersion and Believability*. PhD thesis, Department of Computer and Systems Sciences, Stockholm University. Stockholm, Sweden.

Lankoski, P, Björk, S. 2007. Gameplay design patterns for believable non-player characters. *Situated Play, Proceedings of DiGRA 2007 Conference.* Tokyo, Japan.

Lankoski, P, Björk, S. 2008. Character-Driven Game Design: Characters, Conflict, and Gameplay. *The Sixth International Conference in Game Design and Technology.* Liverpool, United Kingdom.

Loyall, A. B. *Believable Agents: Building Interactive Personalities*. PhD thesis, School of Computer Science, Carnegie Mellon University. Pittsburgh, Pennsylvania, USA.

McGrenere, J., and Ho, W., 2000. Affordances: Clarifying and evolving a concept. *Graphics Interface*, vol. 2000, 179–186.

Warpefelt, H. 2013. *Mind the gap – Exploring the social capability of non-player characters.* Licentiate thesis, Department of Computer and Systems Sciences, Stockholm University. Stockholm, Sweden.

Warpefelt, H. 2015. Cues and Insinuations: Indicating affordances of non-player character using visual indicators. *Proceedings of DiGRA 2015: Diversity of Play.* Digital Games Research Association. Lüneburg, Germany.

Warpefelt, H., Johansson, M., and Verhagen, H., 2013. Analyzing the believability of game character behavior using the game agent matrix. *Proceedings of the sixth bi-annual conference of the Digital Games Research Association: Defragging Game Studies (DiGRA 2013),* Digital Games Research Association. Atlanta, Georgia, USA.

ISBN: 978-989-8533-38-8 © 2015

AUTOMATIC GENERATION OF 3D BUILDING MODELS BY BUILDING POLYGONS ORTHOGONALIZATION

Kenichi Sugihara[1] and Zhen-Jiang Shen[2]
[1]*Gifu Keizai University, 5-50 Kitagata-chou Ogaki-city Gifu-Pref. 503-8550, Japan*
[2]*Kanazawa University, Kakuma-chou Kanazawa-city Ishikawa-Pref. 920-1192, Japan*

ABSTRACT

A 3D city model is important in urban planning and game industries. But, enormous time and labor has to be consumed to create these 3D models. In order to automate laborious steps, a GIS (Geographic Information System) and CG integrated system is proposed for automatically generating 3D building models, based on building polygons on a digital map. Since technicians are drawing these polygons manually with digitizers, depending on aerial photos, not all building polygons are precisely orthogonal. However, creating 3D building models are expected to be orthogonal. When placing a set of boxes as building bodies for creating the buildings, there may be gaps or overlaps between these boxes if building polygons are not precisely orthogonal. In this paper, the new methodology is proposed for rectifying and orthogonalizing the shape of building polygons and generating 3D building models without any gap and overlap.

KEYWORDS

3D city model, automatic generation, GIS (Geographic Information System), 3D building model, Building Polygon, Rectification

1. INTRODUCTION

A 3D city model, such as the one shown in Figure 1 right, is important in urban planning and gaming industries. Urban planners may draw the maps for sustainable development. 3D city models based on these maps are quite effective in understanding what if this alternative plan is realized. However, enormous time and labor has to be consumed to create these 3D models, using 3D modelling software such as 3ds Max or SketchUp. In order to automate laborious steps, we proposed a GIS (Geographic Information System) and CG integrated system for automatically generating 3D building models, based on building polygons or building footprints on a digital map shown in Figure 1 left.

A complicated orthogonal polygon can be partitioned into a set of rectangles. The proposed integrated system partitions orthogonal building polygons into a set of rectangles and places rectangular roofs and box-shaped building bodies on these rectangles. In order to partition an orthogonal polygon, we also proposed a useful polygon expression (RL expression: edges' Right & Left turns expression) and a partitioning scheme for deciding from which vertex a dividing line (DL) is drawn.

Since technicians are drawing building polygons manually with digitizers, depending on aerial photos or satellite imagery as shown in Figure 1 left, not all building polygons are precisely orthogonal. When placing a set of boxes as building bodies for forming the buildings, there may be gaps or overlaps between these boxes if building polygons are not strictly orthogonal. Our contribution is the new methodology for rectifying and orthogonalizing the shape of building polygons and constructing 3D building models without any gap and overlap. In our proposal, after approximately orthogonal building polygons are partitioned and rectified into a set of mutually orthogonal rectangles, each rectangle knows which rectangle is adjacent to and which edge of the rectangle is adjacent to, which will avoid unwanted intersection of windows and doors when building bodies combined.

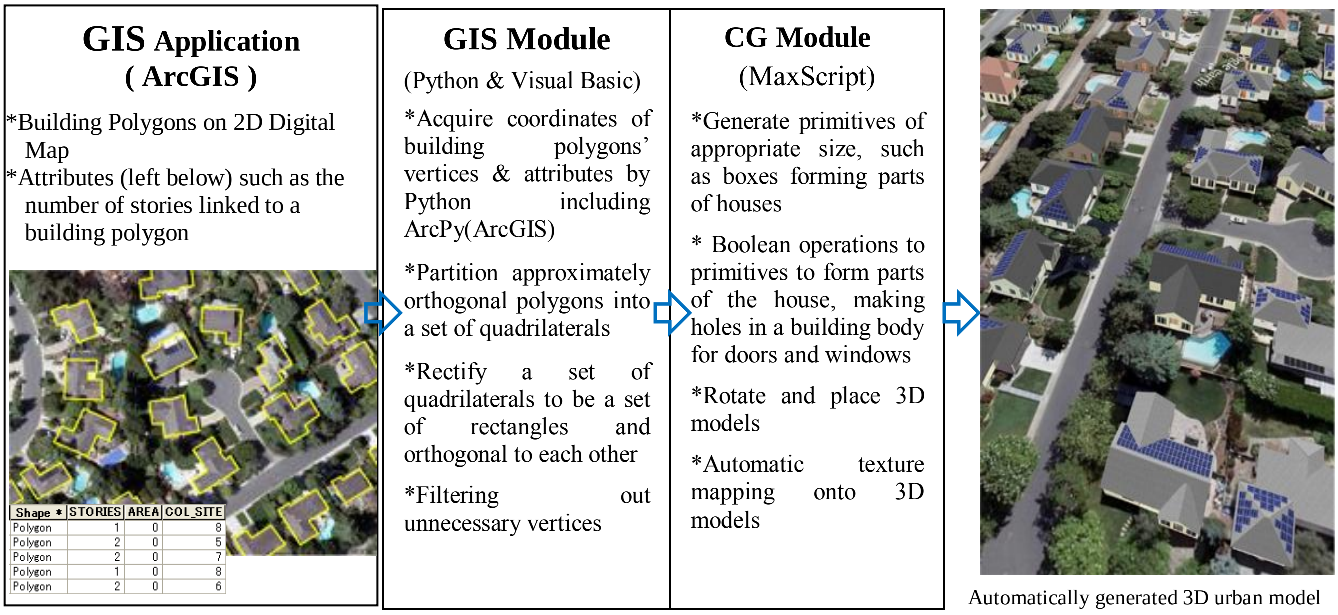


Figure 1. Pipeline of Automatic Generation for 3D Building Models

2. RELATED WORK

Since 3D urban models are important information infrastructure that can be utilized in several fields, the researches on creations of 3D urban models are in full swing. Various types of technologies, ranging from computer vision, computer graphics, photogrammetry, and remote sensing, have been proposed and developed for creating 3D urban models.

Procedural modelling is an effective technique to create 3D models from sets of rules such as L-systems, fractals, and generative modelling language [Parish 2001]. Müller et al. [2006] have created an archaeological site of Pompeii and a suburbia model of Beverly Hills by using a shape grammar that provides a computational approach to the generation of designs. They import data from a GIS database and try to classify imported mass models as basic shapes in their shape vocabulary. If this is not possible, they use a general extruded footprint together with a general roof obtained by the straight skeleton computation defined by a continuous shrinking process [Aichholzer 1995]. By using the straight skeleton, Kelly et al. [2011] present a user interface for the exterior of architectural models to interactively specify procedural extrusions, a sweep plane algorithm to compute a two-manifold architectural surface.

More recently, image-based capturing and rendering techniques, together with procedural modelling approaches, have been developed that allow buildings to be quickly generated and rendered realistically at interactive rates. Bekins et al. [Daniel 2005] exploit building features taken from real-world capture scenes. Their interactive system subdivides and groups the features into feature regions that can be rearranged to texture a new model in the style of the original. The redundancy found in architecture is used to derive procedural rules describing the organization of the original building, which can then be used to automate the subdivision and texturing of a new building. This redundancy can also be used to automatically fill occluded and poorly sampled areas of the image set.

Aliaga et al. [Daniel 2007] extend the technique to inverse procedural modelling of buildings and they describe how to use an extracted repertoire of building grammars to facilitate the visualization and modification of architectural structures. They present an interactive system that enables both creating new buildings in the style of others and modifying existing buildings in a quick manner.

Vanega et al. [Carlos 2010] interactively reconstruct 3D building models with the grammar for representing changes in building geometry that approximately follow the Manhattan-world (MW) assumption which states there is a predominance of three mutually orthogonal directions in the scene. They say automatic approaches using laser-scans or LIDAR data, combined with aerial imagery or ground-level images, suffering from one or all of low-resolution sampling, robustness, and missing surfaces. One way to improve quality or automation is to incorporate assumptions about the buildings such as MW assumption.

ISBN: 978-989-8533-38-8 © 2015

By these interactive modelling, 3D building models with plausible detailed façade can be achieved. However, the limitation of these modelling is the large amount of user interaction involved [Nianjuan 2009]. When creating 3D urban models for urban planning or facilitating public involvement, 3D urban models should cover lots of citizens' and stakeholders' buildings involved. This means that it will take an enormous time and labor to model a 3D urban model with hundreds of building.

Thus, the GIS and CG integrated system that automatically generates 3D urban models immediately is proposed, and the generated 3D building models that constitute 3D urban models are approximate geometric 3D building models that citizens and stakeholder can recognize as their future residence or real-world buildings.

Jianxiong [2014] presents a 3D reconstruction and visualization system to automatically produce clean and well-regularized texture-mapped 3D models for large indoor scenes, from ground-level photographs and 3D laser points. The key component is a new algorithm called "Inverse CSG" for reconstructing a scene in a Constructive Solid Geometry (CSG) representation consisting of volumetric primitives, which imposes regularization constraints to exploit structural regularities. However, with the lack of ground-truth data preventing them from conducting quantitative reconstruction accuracy evaluations, they have to manually overlay their model with a floor plan image.

This computer vision methodology, together with following the Manhattan-world (MW) assumption [Carlos 2010], they use 3D point cloud from laser-scans, extracting line segments passing through them by Hough transformation for structural regularities. In our approach, we only use the position of the vertices of a building polygon for structural regularities or rectification, which reduces the heavy burden of dealing with a huge amount of 3D point cloud.

3. PIPELINE OF AUTOMATIC GENERATION

As shown in Figure 1, the proposed automatic building generation system consists of GIS application (ArcGIS, ESRI Inc.), GIS module and CG module. The source of the 3D urban model is a digital residential map that contains building polygons linked with attributes data shown in Figure 1 left bellow, consist of the number of storeys, the image code of roof, wall and the type of roof (gable roof, hipped roof, gambrel roof, mansard roof, temple roof and so forth). The maps are then preprocessed at the GIS module, and the CG module finally generates the 3D urban model. As a GIS module, a Python program including ArcPy(ArcGIS) acquires coordinates of building polygons' vertices and attributes. Preprocessing at the GIS module includes the procedures as follows:

(1) Filter out an unnecessary vertex whose internal angle is almost 180 degrees. **(2)** Partition or separate approximately orthogonal polygons into a set of quadrilaterals. **(3)** Generate inside contours by straight skeleton computation for placing doors, windows, fences and shop façades which are setback from the original building polygon. **(4)** Rectify a set of quadrilaterals to be a set of rectangles and orthogonal to each other. **(5)** Export the coordinates of polygons' vertices, the length, width and height of the partitioned rectangle, and attributes of buildings.

The CG module receives the pre-processed data that the GIS module exports, generating 3D building models. In GIS module, the system measures the length and inclination of the edges of the partitioned rectangle. The CG module generates a box of the length and width, measured in GIS module. In case of modelling a building with roofs, the CG module follows these steps:

(1) Generate primitives of appropriate size, such as boxes, prisms or polyhedra that will form the various parts of the house.

(2) Boolean operations applied to these primitives to form the shapes of parts of the house, for examples, making holes in a building body for doors and windows, making trapezoidal roof boards for a hipped roof and a temple roof.

(3) Rotate parts of the house according to the inclination of the partitioned rectangle.

(4) Place parts of the house.

(5) Texture mapping onto these parts according to the attribute received.

(6) Copy the 2nd floor to form the 3rd floor or more in case of building higher than 3 stories.

CG module has been developed using Maxscript that controls 3D CG software (3ds MAX, Autodesk Inc).

4. FUNCTIONALITY OF GIS MODULE

4.1 Polygon Expression and Partitioning Scheme

At map production companies, technicians are drawing building polygons manually with digitizers, depending on aerial photos or satellite imagery as shown in Figure 1 left. This aerial photo and digital map also show that most building polygons are approximately orthogonal polygons. An approximately orthogonal polygon can be replaced by a combination of rectangles. When the vertices of a polygon are numbered in clockwise order and edges of a polygon are followed clockwise, an edge turns to the right or to the left by 90 degrees. It is possible to assume that an orthogonal polygon can be expressed as a set of its edges' turning direction; an edge turning to the 'Right' or to the 'Left'.

We proposed a useful polygon expression (RL expression: edges' Right and Left turns expression) for specifying the shape pattern of an orthogonal polygon. For example, an orthogonal polygon with 10 vertices shown in Figure 2 left is expressed as a set of its edges' turning direction; RLRRRRLLRR where R and L mean a change of an edge's direction to the right and to the left, respectively.

The more vertices a polygon has, the more partitioning scheme a polygon has, since the interior angle of a 'L' vertex is 270 degrees and two DLs (dividing lines) can be drawn from a L vertex. In the polygon shown in Figure 2, there are three L vertices, so six possible DLs can be drawn from each L vertex. Among possible DLs, the DL that satisfies the following conditions is selected for partitioning.

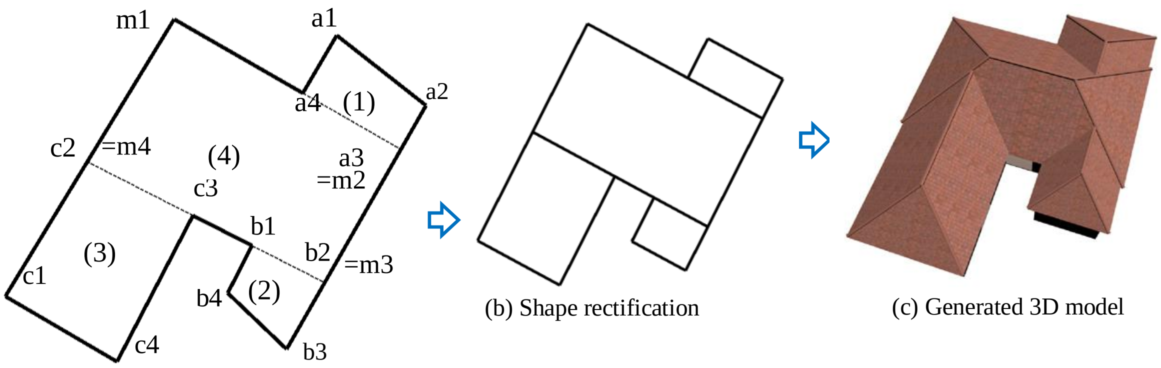
(1) *A DL that cuts off 'one rectangle'.*

(2) *A DL whose length is shorter than the width of a 'main roof' that a 'branch roof' is supposed to extend to, where a 'branch roof' is a roof that is cut off by a DL and extends to a main roof.*

How the system is finding 'branches' is as follows. The system counts the number of consecutive 'R' vertices (=n_R) between 'L' vertices. If n_R is two or more, then it can be a branch. One or two DLs can be drawn from 'L' vertex in a clockwise or counter-clockwise direction, depending on the length of the adjacent edges of 'L' vertex. How the DL is drawn (clockwise or counter-clockwise), that is, 'dividing pattern' is used for reconstructing a rectified polygon and saved at the divided rectangle.

4.2 Active Branch Rectangle

A Figure 2 shows the process of polygon partition and rectification of polygon shape, generation of 3D model. In Figure 2 (a), partitioned rectangles are numbered according to the order of partitioning. For example, rectangle (rect for short) (1) is the branch rect (rectangle) that is partitioned first and will extend to its main roof as shown in Figure 2 (c), automatically generated 3D model.



(a) Polygon partitioning
RLRRRRLLRR

(b) Shape rectification

(c) Generated 3D model

Figure 2. Process of polygon partition and shape rectification, generation of 3D model

In Figure 2 (a), at the time when rect (1) is divided, it is impossible for the rect (1) to know which rect is adjacent to and which edge of the rect is adjacent to. It is not until rect (4) is divided that the rect (1) knows In Figure 2 (a), at the time when rect (1) is divided, it is impossible for the rect (1) to know which rect is adjacent to and which edge of the rect is adjacent to. It is not until rect (4) is divided that the rect (1) knows which rect is adjacent to. After polygon partitioning, the branch rect such as (1) will start to search for an adjacent rect. Rect (2), (3) also has to look for 'one' adjacent rect. Branch rects such as (1), (2) and (3), defined as active rects, will search for an adjacent rect, and will rectify their shape according to the orthogonality. After rectification, divided precise rectangles become orthogonal to each other as shown in Figure 2 (c).

Figure 2 also shows how branch rects are cut off: three dividing patterns. For example, in Figure 2(a), rect (2) is divided by Backward Dividing Line (BDL) which is in the opposite direction in terms of polygon vertices numbering. In Figure 2(a), rect (1) and (3) are divided by Forward Dividing Line (FDL) which is in the clockwise direction.

This dividing pattern is saved at the branch rect, and is used for shape rectification. While these divided branch rects are looking for an adjacent rect and will extend to the adjacent rect. The divided rect whose n_R is more than three can be an absorbing rect, where n_R is the number of consecutive 'R' vertices. After rect (1),(2) in Figure 2 (a) being cut off, rect (4) has three consecutive 'R' vertices (=n_R).

4.3 Shape Rectification

Specifically, the rectification procedure is implemented to the polygon shown in Figure 2 as follows. Before polygon partitioning, all edge length and edge inclination of the polygon are calculated, and the length of all edges are sum up according to the snapped angle of all edge inclination. Then, the angle for a longest sum up edge length can be adopted as the 'main angle' of the polygon, which will be used as the inclination of all partitioned rectangles. Figure 2 shows the process of polygon partition and shape rectification, automatic 3D modelling. A partitioned polygon, i.e., a quadrilateral is numbered clockwise with the start point of a longest edge facing right as pt1 (a1, b1,..) or with the start point of a longest edge facing left as pt3 (a3, b3,..) as shown in Figure 2(a).

When a quadrilateral ('quad' for short) is cut off, the dividing pattern and which edge of the quad is cut off (active edge) is saved at the quad. During the searching stage, an active quad will search for an adjacent quad by locating which quad the checking point on the active edge contains, and then checking on which edge of the adjacent quad the checking point is. In case of quad (1) in Figure 2(a), edge a3a4 will be an active edge, and search for an adjacent quad. After searching and having found out the adjacent quad is quad (4) and the adjacent edge is m1m2, the mutual vertex is a3(=m2), which the rectification procedure uses as a 'standard position (generatrix)' for rectification.

The rectified positions of the vertices of quad (1) are calculated as follows.

$$a1.x=m2.x + w\_S*\cos\theta - w\_L*\sin\theta$$
$$a1.y=m2.y + w\_S*\sin\theta + w\_L*\cos\theta$$
$$a2.x=m2.x + w\_S*\cos\theta : a2.y=m2.y + w\_S*\sin\theta$$
$$a4.x=m2.x - w\_L*\sin\theta : a4.y=m2.y + w\_L*\cos\theta$$

where θ is the main angle and w_S is the average length of two short sides of the rectangle, and w_L is the average length of two long sides of the rectangle. In case of quad (3), the mutual vertex is c2(=m4), which the rectification procedure also uses as a standard position for rectification. The rectified positions of the vertices of quad (3) are calculated as follows.

$$c1.x=m4.x - w\_L*\cos\theta : c1.y=m4.y - w\_L*\sin\theta$$
$$c3.x=m4.x + w\_S*\sin\theta : c3.y=m4.y - w\_S*\cos\theta$$
$$c4.x=m4.x - w\_L*\cos\theta + w\_S*\sin\theta$$
$$c4.y=m4.y - w\_L*\sin\theta - w\_S*\cos\theta$$

The rectified positions of the vertices of a branch quad in other cases are calculated likewise according to the dividing pattern and which edge of the branch quad is cut off.

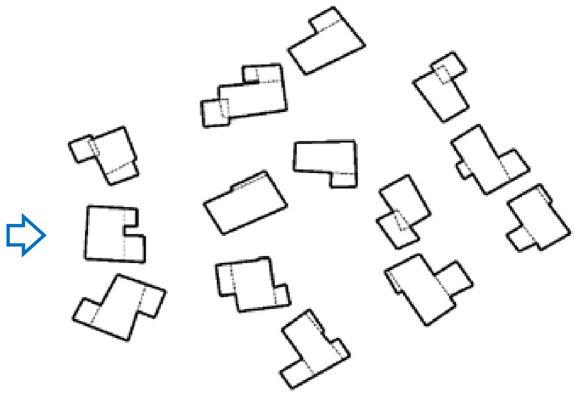
5. CONCLUSIONS AND FUTURE WORKS

Here are the examples of rectification of approximately orthogonal building polygons and an automatically generated 3D urban model in Figure 3. At map production companies, technicians are drawing building polygons manually with digitizers, depending on aerial photos or satellite imagery as shown in Figure 3(a), not all building polygons are precisely orthogonal. However, creating 3D building models are expected to be orthogonal. When placing a set of boxes as building bodies for creating the buildings, there may be gaps or overlaps between these boxes if building polygons are not precisely orthogonal.

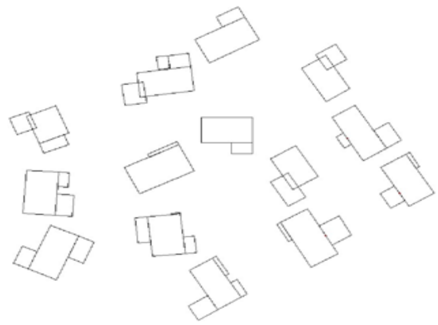
In our proposal, the new methodology is proposed for rectifying the shape of building polygons and generating 3D building models without any gap and overlap. After approximately orthogonal building polygons are partitioned into a set of mutually orthogonal rectangles, each rectangle knows which rectangle is adjacent to and which edge of the rectangle is adjacent to, which will avoid unwanted intersection of windows and doors when building bodies combined for automatic generation of 3D building models.

Our future work includes extension of geometric primitive types to more shapes, such as elliptic cylinders or spheres.

(a) Approximately orthogonal building polygons on satellite imagery



(b) Approximately orthogonal polygon partitioned into a set of rectangles



(c) Active rectangles search for adjacent rectangles, and shape rectification by dividing pattern, depending on mutual orthogonality

(d) Automatically generated 3D urban model

Figure 3. Application of the System: Polygon partition and shape rectification, generation of 3D urban model

REFERENCES

Carlos, V. A., Daniel, A. G., and Bedřich, B. (2010). "Building reconstruction using Manhattan-world grammars", *Computer Vision and Pattern Recognition* (CVPR), 2010 IEEE Conference on: 358 – 365

Daniel, B. R., and Daniel, A. G. (2005). "Build-by-number: rearranging the real world to visualize novel architectural spaces", Visualization, 2005. *VIS 05. IEEE*, 143 – 150

Daniel, A. G., Paul, R. A., and Daniel, B. R. (2007). "Style Grammars for interactive Visualization of Architecture", Visualization and Computer Graphics, *IEEE Transactions on Volume*: 13, 786 – 797

Jianxiong Xiao and Yasutaka Furukawa (2014), "Reconstructing the World's Museums", *International Journal of Computer Vision* (IJCV)

Müller, P., Wonka, P., Haegler, S., Ulmer, A., and Van Gool, L. (2006). "Procedural modeling of buildings", *ACM Transactions on Graphic*s, 25, 3, 614–623

Nianjuan, J. P. T., and Loong-Fah, C. (2009). "Symmetric architecture modeling with a single image", *ACM Transactions on Graphics* - TOG, vol. 28, no. 5

O. Aichholzer, F. Aurenhammer, and D. Alberts, B. Gärtner (1995). "A novel type of skeleton for polygons", *Journal of Universal Computer Science*, 1 (12): 752–761.

Parish, I. H. Y., and Müller, P. (2001). "Procedural modeling of cities", *Proceedings of ACM SIGGRAPH 2001*, ACM Press, E. Fiume, Ed., New York, 301–308

Tom Kelly, Peter Wonka (2011). "Interactive Architectural Modeling with Procedural Extrusions", ACM Transactions on Graphics (TOG) 30 (2), 14, 2011. 41, 2011.

PATH CACHING IN REAL-TIME STRATEGY GAMES

Ignace Pieter Saenen, Jonas El Sayeh Khalil, Peter Lambert and Rik Van de Walle
Ghent University - iMinds - Multimedia Lab, G. Crommenlaan 8 bus 201, B-9050 Ledeberg - Ghent, Belgium

ABSTRACT

This paper proposes a performance optimization for search-based path-planning simulations with the aim to improve systemic scalability. Instead of clustering path requests per agent group, paths are cached at path-request time. Paths are cloned or re-used after selection based on request characteristics, agent properties and external matching criteria. Graph search effort is reduced proportional to the number of agents with similar nearby destinations, while the unique navigational behavior of each individual agent remains unchanged and intact. Formation coherence is maximal for homogeneous navigation, and agent response time improves significantly for large agent groups compared to solutions without this optimization. Explicit multi-agent consensus models are not required and behavioral discontinuities are avoided.

KEYWORDS

Caching, Search, Formations, RTS, Response Time

1. INTRODUCTION

Best-first search algorithms are the enabling technology of Real-Time Strategy (RTS) computer games. In these large dynamic virtual environments, multiple agents concurrently navigate and perform tasks. To navigate, agents frequently request new paths to new destinations. By design, an RTS game challenges the player to micromanage navigating agents into a tactical advantage over a virtual or real opponent by offering multiple levels of fine-grained tactical control. Specific agents have to be steered across a virtual terrain to specific targets to perform specific actions. The predictability of agent behavior during navigation makes the player feel in control at all times, and is crucial for a good game experience. This implies that formation movement of large groups of agents requires rich control, and navigation optimizations must preserve unique agent behavior at all times and with a constant level of quality and performance.

Recent advancements of graph-search algorithms (Abd Algfoor et al., 2015) in 2D grid and lattice space representations (Sturtevant, 2012) and the use of potential fields as described by Koren and Borenstein (1991) and moer recently for the GPU by Demeulemeester et al. (2011) or finite-element based methods (Erra et al., 2004) have allowed RTS-games to upscale the number of concurrently simulated agents. This paper focuses on the scaling performance of mass-agent virtual world navigation simulations, with the goal to improve the response time for concurrent autonomous mobile agents when they move in formation, so that more agents can be simulated in larger dynamic environments, or the simulation becomes feasible for platforms with limited resources.

RTS games are often criticized for exhibiting poor navigational behavior. One of the reasons is that the path-finding system itself can become a bottleneck and will postpone heavy computations – a technique called *time-slicing* - when it fails to honor the real-time constraint. While effective in high-latency environments, this practice becomes problematic when multiple lengthy paths need to be computed at the same time for multiple agents. A burst of concurrent requests then leads to delays, stuttering or erratic results. These problems impair the strategic game design freedom of RTS games.

Path-finding is also used during *local obstacle avoidance* per agent, where the goal is to avoid a collision with an unexpected obstacle by plotting a local course around it, eventually leading the agent back onto its former *global* path. Each navigating agent thus repeatedly composes and solves smaller path requests to guarantee path correctness to the steering module of the agent in order to avoid collisions. Smaller time intervals between local path evaluations make the agent more reactive to dynamic obstacle configurations and repositioning from external forces, but also increase the load on the path solver.

We make three observations for multi-agent path-finding that can be exploited for optimization:

1. A computed path has a high similarity to paths computed from previous requests if the requesting agent and his goal state remain unchanged. The paths will contain mostly the same states and in the same order.
2. Paths of agents with nearby starting and goal states frequently share subsets of the same states in the same order. When two or more agents successfully solve path requests that have both start states and both goal states within a proximity threshold, then there is a probability that computed paths have a high shape similarity or are *isometric*. The probability increases if there are less relevant terrain obstacles in the search space, and if the proximity threshold is smaller.
3. Agents traversing similar paths relatively close to each other in the time domain can be considered to move as an *informal group*.

Spatial cohesion – keeping inter-agent distances and orientations constant - is important if the behavior has strategic or tactical implications. We aim to exploit inter- and intra-agent path similarities and enhance navigation coherence in agent groups by caching and re-using paths at *path-request time*, and show how this can reduce the load on the path. To maximize cache efficiency, priority is given to the computation of a path for a *virtual agent* representing the agent group.

The remainder of the paper is organized as follows: in section 2 we briefly introduce the most relevant existing graph search optimizations based on caching and formation-based navigation research. Section 3 outlines our approach. In section 4 we discuss our results when path-caching is applied to Dragon Commander, a commercial RTS game by Larian Studios. We end with conclusions in section 5.

2. RELATED WORK

2.1 Graph Search

Caching of (sub-) paths is used to optimize node expansion in graph search algorithms. We discuss the most relevant approaches: In the class of hierarchical path-finding algorithms (Botea, 2004, Jansen and Buro, 2007, Caggianese and Erra, 2014 and Bnaya & Felner, 2014) the state graph is transformed to a higher-order state graph that has all possible paths pre-computed. The final path is then computed between higher-order states; Fringe-search (Björnsson, et al., 2005) is based on memory-enhanced Iterative Deepening (Plaat et al. 1995) and is ideal for large search spaces. Here, the states in the open-list are cached and serve to jump-start the next IDA* expansion iteration; Botea (2011) brute-forces path pre-computation. All (optimal) paths between all state pairs are compressed into a Compressed Path Database (CPD), and run-time path-finding is reduced to a look-up. Our proposed caching solution is coextensive to these methods.

2.2 Formations

Cooperative behavior for multi-agent navigation is described using a *consensus* model. It describes formation leadership using inter-agent communication protocols and hierarchical control structures (Olfati-Saber, 2006). Optimizations relevant to path-finding remain within the hierarchical bounds of the implemented model, and such explicit control structures scale linearly or exponentially when the number of simulated agents is increased.

Conflict resolution - assigning navigation tasks to agents - is handled within the formation using an evaluation function, e.g. based on a potential field or group-torque analysis (Li and Chen, 2005 and Chen and Li, 2006). To identify the relevant agents, a form of K-means clustering, hierarchical clustering or ant-based clustering (Handl and Meyer, 2007) manages the control hierarchy, e.g. when logically grouped agents are dispersed across the state-space. Static clustering is not well suited for highly dynamic situations and is difficult to align with the design requirements of an RTS game. Instead, our approach clusters agents at path-request time.

Transfer of ownership of the leader-role between agents causes navigational discontinuities. In highly dynamic scenarios, agents are quickly entering and leaving the simulation. Hence, the position of the agent bestowed with the leader-role can change rapidly. If the leading agent communicates paths for other agents to follow, these following agents will exhibit the same control discontinuities during navigation, leading to frequent formation restructuring and in-consequent planning on the agent level.

To our knowledge, existing solutions prioritize formation structure over individual agent behavior, be it implicitly or in an explicit top-down approach. This is adverse to the design goals of an RTS game, where every agent has specific properties and behavior that represent their tactical purpose. A solution should therefore not adapt agent behavior when formations encounter dynamic obstacles, or when the control structure itself changes. To that end, our method is agnostic to – but compatible with - existing agent control mechanisms if present.

3. METHODOLOGY

Without loss of generality, we assume that one path- finding system is used for all path computations, and all searches terminate. A maximum to the number of search expansions is not imposed, although we used it to emulate low-spec hardware during testing.

Path acquisition for agent group $G \subseteq A$ can be organized in four steps: (1) path formalization, (2) path-matching, (3) path evaluation and caching, and (4) path retrieval. These steps operate in an abstraction layer between the agents and the graph-search algorithm, as illustrated in Figure 1, and are discussed in the next four sections below.

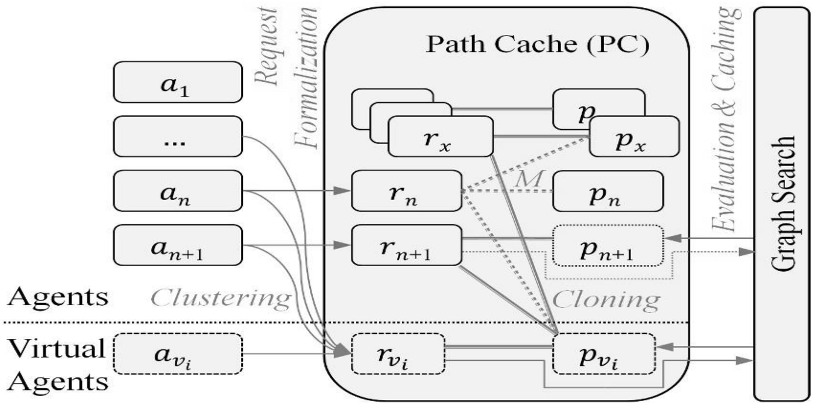


Figure 1. Similar path requests are identified during request matching and cached in path cache PC. Graph search is used when matching criteria M fail to find a matching path p_i for request r_i. Formation coherence improves when virtual agent path p_v is recycled.

The following abbreviations are briefly introduced so that the proposed caching and matching strategy can be formalized. For agent $a_i \in$ the set of all agents A, path request r_i in state-space S from starting state s_i to goal state g_i has optimal path $p_{i,0}$ out of the set of valid paths P_i for r_i. Path $p_{i,0}$ is an ordered set of l states n_j such that $p_{i,0} = \{s_i = n_0, n_1, .., n_{l-2}, n_{l-1} = g_i\}$ and $n_j \in S, j \in [0..l[$.

3.1 Path Formalization

When goal state g_i - a new destination location - is assigned to agent $a_i \in G$, a search request r_i is formalized and describes agent a_i and states s_i and g_i.

Search request r_i can be optionally extended with a group descriptor G_d that expresses the *spatial coherence* in G, and consist of a centroid h_c and size h_s so that $G_d = \{h_c, h_s\}$. In our implementation, both are derived from incremental convex hull algorithm by Melkman (1987). $\forall a_i \in G$: all G_d are equal.

Virtual agent a_v is created from G_d if and only if G_d is proportional to the sum of all n agent sizes. This enables the construction of virtual path p_v for all agents in G. Otherwise, G_d is removed from request r_i.

3.2 Path Matching

Path cache PC is queried for the path specified by r_i using a matching operator M based on a set of *matching criteria*. These criteria include proximity thresholds for matching start states and goal states, but can include additional criteria such as collision masks and avoidance masks.

If one or more paths $\{pc_{i,j} \mid j \in [1..k]\}$ fulfill the matching criteria of M, a *winner-takes-all* strategy is used to select and mark the most relevant path $pc_{i,r}$, e.g. based on path complexity. This path $pc_{i,r}$ is subsequently cloned for the requesting agent a_i. Otherwise, the request r_i remains unresolved.

If r_i contains group descriptor G_d, then M simulates the virtual group agent a_v instead of agent a_i that originated the path request, and uses $s_v = hc$ as starting state and $g_v = g_i - s_i + hc$ as goal state of virtual request r_v during graph search.

3.3 Path Evaluation and Caching

The path cache PC contains both computed paths *and* path requests. This makes it possible to cache and match *requests* that *finish computation in the future*. In that case, all other matching requests will be notified for further evaluation. Short resolved paths with a small number of states are not cached to minimize path-matching overhead. Every agent computes a new path for r_i, or uses a cloned path $pc_{i,r}$ from section 3.2.

Paths are computed using a best-first iterative graph search algorithm operating on a global state-space configuration S (e.g. a 2D grid). When the search algorithm becomes available for computation, it iterates a limited number of times *PSe*.

path p_i is computed within *PSe* iterations, it is considered trivial to calculate and is not added to the path cache PC. Otherwise, the elapsed numbers of iterations augment r_i's potential relevance when re-issued from the PC, therefore r_i is added to the PC. r_i is also cached if it has group descriptor G_d, or if a graph search is busy and the minimum distance between s_i and g_i exceeds a threshold ct.

If computed path p_i has *complexity* $> cn$ (configurable constant), or has path length $> ct$, then p_i is added to the path cache PC.

As an optimization, r_i can be pruned early if a single straight line segment constitutes a valid path from s_i to g_i.

3.4 Path Retrieval

When path $pc_{i,r}$ is retrieved from path cache PC, it is *cloned* and *localized* for requesting agent a_i. This is illustrated in Figure 2 (a). Here, p_i is constructed by duplicating all states in p_j except for the start and goal states, which are substituted by the start and goal states of r_i.

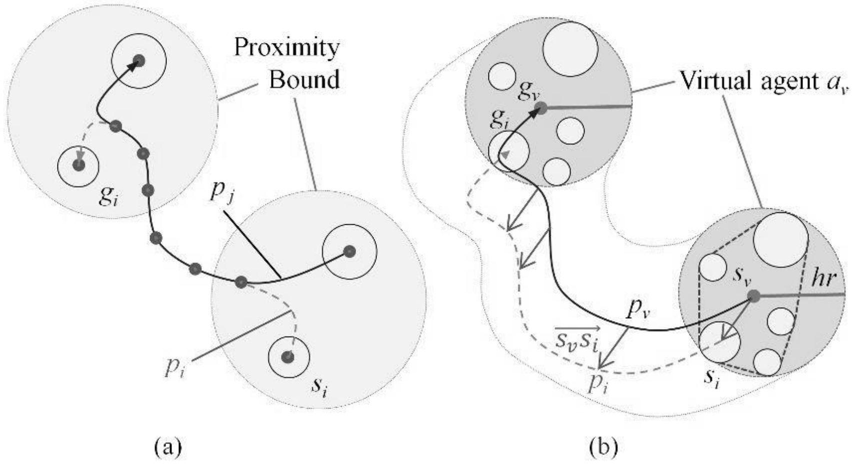


Figure 2. (a) New path p_i is cloned from cached path p_j and localized; (b) New path p_i is derived from virtual path p_v that is constructed for virtual agent a_v.

If group descriptor G_d is valid, then path p_v - computed for virtual agent a_v - is cloned and p_i is constructed by duplicating all states in p_v and offsetting them over vector $s_i s_v$. See Figure 2 (b).

Paths are localized by substituting states s and g for states s_i and g_i, respectively. If the path is valid, it is optimized, e.g. using a smoothing algorithm.

If p_i is invalid, a new path request is created, but any G_d is omitted. If G_d was invalid or not present, the request is excluded from path-matching and path-caching.

3.5 Path Removal

Large path caches impact the efficiency of the path matching step and should be avoided. Many different caching strategies can be used to prune unused paths from the path cache *PC*. This work focuses on the computational worst-case scenario that is established by assuming that no pre-computed path exists in the *PC*. Hence, our implementation removes a path from *PC* if it is no longer in use *and* all cloned copies of it are no longer in use.

On the other hand, *keeping* paths longer in the path cache *PC* is advantageous when the probability for requesting similar paths is high, for example when agents exert a patrolling behavior between two or more *waypoints*. In this case, the path can be simply marked by the application so that it remains in the cache. Although this was not explored further and out of the scope of this work, the balance between cache hit ratio and path computation complexity can be further optimized using methods such as reference counting with ageing, or simulated annealing for successfully cloned paths.

4. EVALUATION

Our approach was applied to the computer game Dragon Commander, a fast-paced RTS game that is played by creating, moving and using tactical combat units (agents) against other (human or artificial) players on various 2D terrain types. Three types of combat units - naval, ground and air units of various sizes and each type with different capabilities – can be created and upgraded using available resources such as gold, population and research points. These resources can be collected when the player exerts control over specific portions of the terrain long enough. A play session ends when one player conquers a predefined area, or disables or destroys all opposing units and buildings.

To evaluate the effects of path caching, a number of groups of ground units - each with a randomly selected size - are sent repeatedly to a set of randomized destinations on the 2D grid. All unit destinations lie within close proximity so that units do not overlap, i.e. with respect to the various sizes of the units. The 2D grid dimensions are 100 x 100 with approximately half the terrain covered by large obstacles such that a planned path is either possible or impossible for *all* unit *sizes*. For testing purposes, the search expansion rate during path computation is constrained to simulate an elevated workload similar to that after a request burst, so that path-caching is enforced.

The average group response time is measured by taking the elapsed search iterations between the group path request and the start of navigation per agent, subtracted by graph search time, and normalized by average path complexity for that request. Figure 3 shows decoupled average response times from graph complexity and group size.

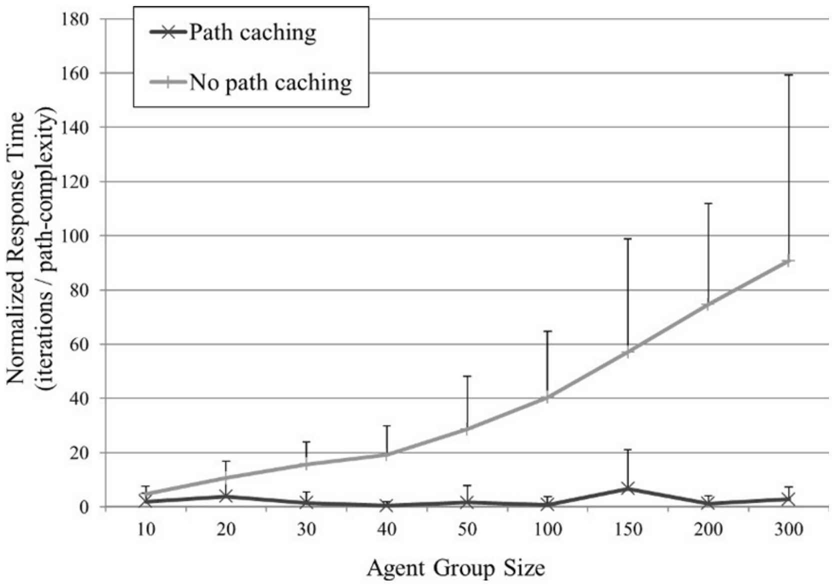


Figure 3. Group response time improves proportional to the number of agents. Fluctuations stem from combinations of group sizes and grid space that trigger different scenarios.

4.1 Scaling

Our proposed solution provides three new types of control to improve systemic scaling.

- The *path-matching operator M* decides if a cached path is cloned based on local and global information. For example, to simulate anxiety or panicking behavior, re-use of paths can be limited so that agents must navigate individual paths and disregard formation. Relaxing or restricting the matching criteria thus reduces or increases the workload of the search algorithm and changes the scalability of the system.
- The *path-caching criteria* determine the caching efficiency of *PC*. For example, when more agents enter the simulation, the caching criteria can be relaxed so that more paths are cached. This type of control at the path-finding level is difficult to reproduce with explicit consensus models that reason about logically isolated groups.
- *Path cloning* implements path optimizations per agent in an enriched context. Each matching request has a group descriptor, or a valid virtual agent path, or is informed that the path is already used by other agents. This information is used to reduce the workload on the graph search, and to improve path coherence. For example, agents in a group can maintain inter-agent distances.

As a result, we observed improved agent reactivity and local obstacle handling due to the higher availability of the graph search algorithm. In best-case scenarios, all agents start navigating as soon as a path is shared.

4.2 Agent Behavior

Agent navigation behavior is defined per individual and thus potentially heterogeneous during execution. Path caching allows agents to reuse paths with high similarity and navigate accordingly. Groups with increasingly homogeneously behaving agents have higher formation *coherence* during navigation and tend to preserve formation *orientation*.

At the same time, agents with heterogeneous velocities or steering patterns are visibly identifiable from their observed behavior and can freely abandon formation coherence during navigation, making them easily recognizable, predictable and controllable.

4.3 Matching Operator M

The operator M iterates over all cached paths in path cache PC and finds the best possible match $pc_{i,r}$ for a new request r_i. In our evaluation, the matching conditions for using or cloning a cached path pc_i based on source state s_i and goal state g_i of each new path request r_i are as follows:

- Path *completeness*: The path pc_i must have been allocated for usage or be in use. Indeed, the allocated path may not have finished computation. Alternatively, if the path has not been allocated for usage, it is scheduled for removal since no agent is using it.
- Agent *validity*: First, the agent requesting the path must have compatible properties compared to the agent that originally provided the path to the path cache. Put differently, agents of different logical search spaces should not be sharing paths, for example when agents are only allowed to move over specific terrain types such as shallow water or desert sand. Secondly, if the original agent allowed the path computation to stop the search early within an arbitrary distance to the goal state, this also must work for the agent that submits the new path request. Finally, the path provides the verified maximum agent size that is able to traverse the path on the map without collisions. The path pc_i is eligible for sharing if the requesting agent size is smaller or equal to the size of the original agent (e.g. virtual agent *av*).
- Path *optimality*: The path pc_i is rejected when distance *d1* between the agent source position and the path source state s_i exceeds a threshold *t1*. Similarly, the path is rejected when the distance *d2* between the agent goal position and the path goal state g_i exceeds a threshold *t2*. Finally, the operator minimizes $\alpha d1 * \beta d2$, and thus prefers paths closest to the requesting agent, where the ratio α/β biases path preference towards the source or destination.

5. CONCLUSION

In this paper, an optimization for loosely coupled heterogeneous agent path planning in real-time simulations is presented. The main contribution of the method is to model and optimize path similarity at the path-finding level, and augment control over the scalability of the system. By exploiting informal group navigation and related path request behavior at request-time, computational complexity can be reduced while retaining individual agent behavior characteristics. The number of concurrent path requests can be controlled by relaxing or tightening path-caching or matching criteria for the path cache. The resulting decoupling of agent response time from the system's agent complexity was demonstrated in a commercial game environment using an implementation of our algorithm.

No preprocessing or modifications to agent behavior are required, and the system is agnostic to the underlying graph-search algorithm and agent consensus model.

ACKNOWLEDGEMENT

This research is funded by Ghent University and iMinds. The authors like to thank Larian Studios for the use of the RTS game Dragon Commander, and Nathan Sturtevant for making many benchmark maps available for search algorithm research related to games.

REFERENCES

Abd Algfoor, Z., Sunar, M. S., & Kolivand, H. (2015). A Comprehensive Study on Pathfinding Techniques for Robotics and Video Games. *International Journal of Computer Games Technology*, Vol. 2015.

Björnsson, Y., Enzenberger, M., Holte, R. C. & Schaeffer, J. (2005). Fringe Search: Beating A* at Pathfinding on Game Maps. *Proceedings of the IEEE Symposium on Computational Intelligence and Games (CIG).* held 4-6 April 2005 at Essex University. Colchester, United Kingdom: IEEE, pp. 125-132.

Bnaya, Z. & Felner, A. (2014). Conflict-Oriented Windowed Hierarchical Cooperative A∗. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).* Hong Kong, People's Republic of China: IEEE, pp. 3743-3748.

Botea, A., Müller, M. & Schaeffer, J. (2004). Near optimal hierarchical path-finding. *Journal of game development (JOGD), 1*. Charles River Media, 7–28.

Botea, A. (2011). Ultra-fast optimal pathfinding without run-time search. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*. Raleigh, North Carolina, USA: AAAI Press

Caggianese, G., & Erra, U. (2014). Parallel hierarchical A* for multi agent-based simulation on the GPU. In *Euro-Par 2013: Parallel Processing Workshops (2014)*, Springer, pp. 513–522.

Chen, X., & Li, Y. (2006). Smooth formation navigation of multiple mobile robots for avoiding moving obstacles. In *International Journal of Control Automation and Systems,* Vol. 4, pp. 466.

Demeulemeester, A., Hollemeersch, C.-F., Mees, P., Pieters, B., Lambert, P., and Van de Walle, R. (2011). Hybrid path planning for massive crowd simulation on the GPU. In *Proceedings of the International Conference Motion in Games*. Springer, pp. 304–315.

Erra, U., De Chiara, R., Scarano, V., & Tatafiore, M., (2004). Massive simulation using GPU of a distributed behavioral model of a flock with obstacle avoidance. In *Proceedings of the 9th International Fall Workshop on Vision, Modeling and Visualization*, Vol. 2004.

Handl, J., & Meyer, B. (2007). Ant-based and swarm-based clustering. *Swarm Intelligence,* Vol. 1, Issue 2, Springer, 2011, pp. 95–113.

Jansen M. R., & Buro M. (2007). Hpa* enhancements. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, AAAI Press, pp. 84–87.

Koren Y., & Borenstein J. (1991). Potential field methods and their inherent limitations for mobile robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, IEEE, pp. 1398–1404.

Li Y., & Chen X. (2005). Leader-formation navigation using dynamic formation patterns. In *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, IEEE, pp. 1494–1499.

Melkman A. A. (1987) On-line construction of the convex hull of a simple polyline. *Information Processing Letters,* Vol. 25, Issue 1, Elsevier, pp. 11–12.

Olfati-Saber R. (2006). Flocking for multi-agent dynamic systems: Algorithms and theory. In *IEEE Transactions on Automatic Control,* Vol. 51, IEEE, pp. 401–420.

Plaat, A., Schaeffer, J., Pijls, W., & De Bruin, A. (1995). Best-first fixed-depth game-tree search in practice. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Canada: AAAI Press, pp. 273–281.

Sturtevant N. R. (2012). Benchmarks for grid-based pathfinding. In *IEEE Transactions on Computational Intelligence and AI in Games*, Vol 4, Issue 2, IEEE, pp. 144–148.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

EXTREME ASSET SIMPLIFICATION AND THE PRESERVATION OF VISUAL APPEARANCE

Jonas El Sayeh Khalil, Ignace Saenen, Peter Lambert
and Rik Van de Walle
Multimedia Lab, Ghent University - iMinds, Belgium

ABSTRACT

Reusing animation film assets for real-time rendering requires extreme simplification. As well-known simplification approaches do not suffice, studios are still forced to manually simplify their assets. To automate this, we employ a pipeline for efficient geometry-based simplification and make use of normal mapping to ensure visual similarity. Our obtained results are promising: geometric complexity is vastly reduced while maintaining a recognizable model, unlike results with classical simplification approaches as employed by commercial applications. We have compared the approaches in two settings, aiming at a similar number of triangles and aiming at a similar storage size, both of which prove that our extreme asset simplification is a valid alternative for classical topological simplification approaches.

KEYWORDS

Real-time rendering, asset simplification, appearance preservation.

1. INTRODUCTION

Computer-generated imagery renders assets nearly indistinguishable from real life; the recent production of *Life of Pi*[Bayever2013] is a great example of this. Highly accurate models, materials, and lighting simulations require several hours of rendering time per frame, a large time-per-frame budget which stands in stark contrast to the 1/60s available in a real-time rendering environment[Akenine2008] for updating and rendering a frame. Considering the current evolution of applications towards mobile devices, even greater restrictions are posed by the rendering hardware, networking capabilities and battery lifetime of such devices.

Reusing high-quality assets for real-time applications thus requires simplification. This not only implies a large reduction of the number of geometric primitives, but also requires reconsidering the use of materials and the composition of models to reduce the number of draw calls and to avoid wasting rendering resources.

There is a dire need for such an extreme asset simplification, as recently several animation film studios started creating game content based on their film assets. *Disney INFINITY*[Disney2014] is an example; with their platform, they are already aiming at mobile gaming [Butts2013]. Today, such endeavors require a manual redesign of the assets to allow interactivity. We propose a step towards automation by reducing complex geometries to very simple models, while maintaining a high level of visual quality by using textural data. This will lead to a more cost-efficient production of games based on existing animation film assets.

1.1 Contributions

In this work, we make use of the conversion pipeline proposed in [Khalil2014] to create a low-resolution approximation of the geometry of a film asset. By applying normal mapping to the generated surfaces, we are able to preserve the visual appearance of an asset. Furthermore, we have made a comparison with well-known simplification approaches as found in both commercial and non-commercial applications, showing that our approach generates pleasing results at a fraction of their number of triangles.

To our knowledge, no solution has been proposed yet for full asset simplification, forcing studios to redesign their assets manually. We propose a first step towards (semi-)automatic simplification, focusing on reconstructing the appearance of the original model, and showing that the results remain visually pleasing.

ISBN: 978-989-8533-38-8 © 2015

2. RELATED WORK

This work focuses on extreme asset simplification, a subject which falls between the domains of mesh processing and modeling. As a simplification approach, it is part of a larger bulk of literature covering mesh simplification discussed in Sec. 2.1. Mesh simplification as such proves to be insufficient considering the order of magnitude and the requirements of the simplification, which demands a remodeling of the assets. To our knowledge, no solution has yet been proposed for automatic *asset* simplification. We discuss several related issues in Sec. 2.2, and describe work tackling two of the issues in Sec. 2.3 and Sec. 2.4.

2.1 Mesh Simplification

Mesh simplification considers the problem of reducing the number of triangles in a single polygonal mesh. Well-known initial efforts are approaches such as the *mesh decimation* approach of Schroeder et al.[Schroeder1992] and the *retiling* approach proposed by Turk[Turk1992]. Furthermore, the *Quadric Error Metrics* described by Garland and Heckbert[Garland1997] is used often. Good overviews of the initial efforts are given by Cignoni et al.[Cignoni1998] and by Talton[Talton2004].

More recently, *simplification* has been intertwined with *progressive representations* and *multiresolution compression*: models are still simplified, but in a lossless manner that has to allow reconstruction, as reflected by state-of-the-art coding systems such as [Valette2009] and [Lee2012]. This reconstructability limits the simplification performance. A recent approach that does *not* focus on progressive representations is the work of Coll and Paradinas [Coll2010]. They described a simplification approach which handles models with multi-chart textures. This is a valid step towards *asset* simplification as it takes into account more than just geometric properties. However, the approach still assumes a single "well-formed" two-manifold mesh, an assumption that cannot in general be made for animation film assets.

Jeff Somers has developed a mesh simplification tool [Somers2015], which implements several simplification approaches. Firstly, a simplified version of the *Quadric Error Metrics* was implemented. Secondly, the polygon reduction algorithm described by Stan Melax [Melax1998] was implemented as well. Finally, a shortest-edge-first collapsing approach was implemented for sake of comparison.

The CGAL Open Source Project [Alliez2008] supports mesh simplification as well [Cacciola2014]. The same issue arises: they only process *oriented 2-manifold surfaces with any number of connected components*. Our simplification approach on the other hand, does not make any assumptions on the connectivity of the high-resolution mesh.

2.2 Asset Simplification

In essence, our work is a first step towards tackling extreme *asset* simplification. An asset is defined here as a virtual 3D model, consisting of a mesh and all of its accompanying data. This includes textural data such as diffuse maps, specular maps and normal maps, animation data for animated assets, material data to represent surface properties, and more. In this sense, extreme asset simplification encompasses more than just a mesh simplification step.

We denote the simplification as *extreme*: the conversion from film-quality assets to game assets requires a downscaling to such an extent that classical mesh simplification approaches are insufficient, even when only considering the simplification of the mesh geometry itself. In general, a mesh repair step is required to ensure two-manifoldness as required by most mesh simplification approaches.

Furthermore, a major issue with animation film assets is their composition consisting of *multiple* meshes. Amongst others, this is done to facilitate asset modification and animation. To illustrate this, an eye can be modeled as having a cornea, an iris, a pupil and an eyeball[Blender2014]. The use of multiple meshes and multiple materials allows artists to, e.g., easily modify the color of the iris or animate the pupil. This level of detail in general does not allow real-time rendering, and a mere reduction of the number of triangles does not suffice. Trade-offs have to be made between geometric and textural representations; in this specific example, the eye could be modeled as a single sphere, with the cornea, iris and pupil modeled together as a simplified (animated) texture mapped on the eyeball, or the eyes could even be modeled through textural data mapped on the geometry of the head.

In this work we consider two specific issues: (1) meshes need to be merged as we want to reduce the number of disjoint meshes to be rendered, and (2) permanently hidden parts should be discarded as we want to avoid drawing parts of surfaces that are never visible.

2.3 Mesh Merging

To obtain a single triangle mesh, multiple disjoint meshes of an asset have to be merged. This problem has been investigated in solid modeling. In *Constructive Solid Geometry (CSG)*, designers create complex models through addition and subtraction of more primitive models. However, the binary space partitioning trees used to define the hierarchical operations [Naylor1990] have a high complexity, making real-time CSG too costly to use for real-time rendering.

The problem of merging disjoint meshes has also been tackled for 3D surface scanning, as each scan is able to capture only part of a model. Turk and Levoy [Turk1994], and later Marras et al.[Marras2010], propose a zippering approach to join meshes where they intersect geometrically. Contrary to these scanned meshes where intersections are found where scanning ended, asset conversion requires the complete meshes to be compared amongst each other to find intersections, making this operation unfeasible for assets consisting of many complex meshes.

2.4 Hidden Surface Removal

When meshes are merged to a single mesh, some parts can be permanently hidden and should not be retained. Topics concerning *hidden surface removal* focus on real-time, view-dependent removal. For asset conversion, all permanently hidden data must be dropped in a *preprocessing* step. The use of ambient occlusion to find and remove hidden triangles has been suggested in a web article concerning practical uses of MeshLab [ALoopingIcon2009], but this is not a perfect solution as visible parts are possibly removed. An alternative is to use a volumetric representation and fill an objects interior, as we suggested in [Khalil2014].

3. APPROACH

We will discuss our pipeline proposed earlier [Khalil2014] in Sec. 3.1, describing the transformation of animation film assets to coarse geometric approximations via voxel models. After a post-processing optimization step as discussed in Sec. 3.2, we generate and use normal maps as described in Sec. 3.3, allowing the preservation of the visual appearance of an original asset.

3.1 Voxelize, Process and Remesh

To avoid complex topological issues such as non-manifoldness and improperly connected or even self-intersecting meshes, intermediate voxel models are used to represent each mesh, resulting in geometric approximations. The voxel resolution determines the size under which holes and handles will be filled; larger ones are preserved. In the case of designer inaccuracies, filling these holes is certainly desirable. If holes or handles are present by design, these will either be recreated *visually* through the use of textural data, or a finer voxel resolution can be chosen to preserve these holes and handles *within the mesh itself*.

In a subsequent step, the multiple meshes of an asset are merged in the voxel-domain, where this merge operation translates to a voxelwise OR-operation. This elegantly combines multiple meshes without requiring complex tests to find where they are intersecting.

The next step removes permanently hidden parts. This is done by assuming that the object will always be viewed from outside of its bounding box. Determining the permanently hidden parts is done by first determining the visible space. This is done by using a flood filling algorithm starting from a corner voxel outside of the model, i.e., a voxel on the border of the bounding box. Voxels that have not been filled therefore must be permanently hidden and can safely be marked as part of the model.

Finally, the voxel model is transformed back to a triangle mesh using the Marching Cubes algorithm [Lorensen1987]. As all permanently hidden parts have by now been marked as part of the model’s interior,

ISBN: 978-989-8533-38-8 © 2015

Marching Cubes will not generate triangles that are never visible; this avoids wasting precious rendering resources. Intersecting models are properly handled, as the voxel model does not show any distinctions between meshes anymore.

3.2 Merge Coplanar Faces

The Marching Cubes algorithm is known for generating large numbers of triangles. We addressed this by merging coplanar faces and retriangulating them, resulting in a large decrease in the number of triangles without altering the geometry. Any simplification algorithm can be considered at this step of the pipeline: as all topological issues have been resolved, simplification at this point will have superior results compared to mesh simplification *before* voxelization. This can be combined with a higher voxel resolution to increase the resulting geometrical quality if this is desired. However, here we have not implemented such a simplification post-processing step as we want to illustrate the resulting quality using relatively low voxel resolutions.

Note that we have been careful not to change the positions of the remaining vertices, i.e., they are a subset of the original vertices. Due to the regular sampling during the voxelization process, this allows us to represent the resulting vertices using a limited number of bits, as they are implicitly quantized.

3.3 Generate Normal Maps

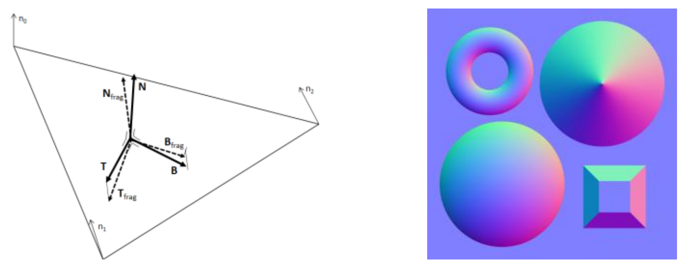


Figure 1. (left) Example normal basis. The bold full lines show a triangle-wide normal basis, which is subsequently refined to the fragment-specific basis in dashed lines. (right) Classical example of a tangential space normal map

The loss in visual detail can in practice be compensated through the use of textural data: (1) normal maps define the corrected orientation of the normal across the model surface, and (2) displacement maps define the desired local displacement of the surface. Whereas normal maps only influence the lighting equations, displacement maps support on-the-fly geometry generation.

Currently we have investigated the effect of using only normal mapping as these are straightforward to implement on mobile devices, which is a main target of our asset simplification. We have generated tangent space normal maps using Autodesk Mudbox. The visualization however has been implemented in our custom renderer for ease of comparison, as described next.

3.3.1 Render Resulting Models

For tangent space normal maps, there are still several ways to interpret the color values within the image. Depending on whether or not the normal component can be negative, or how the tangential axes are oriented, their interpretation can change. Based on a whitepaper by Autodesk[Autodesk2011] we were able to reconstruct the correct use of the normal maps generated by Mudbox. We define (**T**, **B**, **N**) as the tangential basis with **T** the tangential axis, **B** the bitangential axis, and **N** the normal axis (shown in bold in Figure 1). Within each triangle this basis is constant, and is transformed to object space using following equations.

Calculations in the geometry shader. In the geometry shader, we find a tangential basis for the entire triangle. With $\mathbf{p}_i$ the position of the i^{th} triangle vertex and (u_i,v_i) its texture mapping coordinates, the axes **N, T** and **B** are defined as follows:

$$\mathbf{N} = [(\mathbf{p}_1 - \mathbf{p}_0) \times (\mathbf{p}_2 - \mathbf{p}_0)] / \|(\mathbf{p}_1 - \mathbf{p}_0) \times (\mathbf{p}_2 - \mathbf{p}_0)\| \quad (1)$$

$$s_1 = u_1 - u_0,\ s_2 = u_2 - u_0,\ t_1 = v_1 - v_0,\ t_2 = v_2 - v_0 \quad (2)$$

$$\mathbf{T'} = [t_1(\mathbf{p}_1 - \mathbf{p}_0) - t_2(\mathbf{p}_2 - \mathbf{p}_0)] / [s_2 t_1 - s_1 t_2] \quad (3)$$

$$\mathbf{T} = \mathbf{T'} / \|\mathbf{T'}\| \quad (4)$$

$$\mathbf{B} = \mathbf{N} \times \mathbf{T} \quad (5)$$

Adaptations in the fragment shader. In smooth rendering however, the fragment normal **n** is obtained by interpolating the surrounding vertex normals, denoted n_0, n_1 and n_2 in Fig. 1. Using this normal **n**, each fragment can have a unique transformation of the tangential basis ($\mathbf{T}_{frag}$,$\mathbf{B}_{frag}$,$\mathbf{N}_{frag}$), as shown in dashed lines in Fig. 1. To ensure orthogonality, the Gram-Schmidt process[Hazewinkel2001] is applied (Eq. 7 and 8).

$$\mathbf{N}_{frag} = \mathbf{n} / \|\mathbf{n}\| \tag{6}$$

$$\mathbf{T}_{frag} = \mathbf{T} - (\mathbf{N}_{frag} \cdot \mathbf{T})\, \mathbf{N}_{frag} \tag{7}$$

$$\mathbf{B}_{frag} = \mathbf{B} - (\mathbf{N}_{frag} \cdot \mathbf{B})\, \mathbf{N}_{frag} - (\mathbf{T}_{frag} \cdot \mathbf{B})\, \mathbf{T}_{frag} / \|\, \mathbf{T}_{frag} \,\| \tag{8}$$

Calculation of normals. For each color component r, g and b, the interval [0,1] is transformed to the interval [-1,1], to be used as coordinates in the transformed basis:

$$t = 2r - 1,\ b = 2g - 1,\ n = 2b - 1 \tag{9}$$

$$\mathbf{n}_{final} = t\, \mathbf{T}_{frag} + b\, \mathbf{B}_{frag} + n\, \mathbf{N}_{frag} \tag{10}$$

As the normal components are largely positive, the blue color components are often larger than 0.5, resulting in the well-known bluish color of tangent space normal maps such as depicted in Fig. 1.

4. EVALUATION

For the evaluation of our approach we provide visual results, as objective quality measurement for comparing rendered 3D models is still an unsolved problem. We emphasize that the goal is not exact reconstruction but a rendered image which is visually pleasing and recognizable; many of the converted assets will not play a major role in a video game. Due to the spatial and practical limitations, we provide a limited selection of the visual results in this paper; we offer more results as additional material, showing the results under changing lighting conditions (https://www.dropbox.com/sh/x0bo2ol93g5hto6/AAA_nx5k6Kgwak1o_xnDlGbia).

4.1 Experimental Setup

The experiments have been performed using actual animation film assets provided by Studio 100 NV, showing typical issues such as permanently hidden parts, topological inaccuracies or compositions of hundreds of smaller objects. The assets are stripped of all material and texture information as we focus on geometry simplification in this paper (see additional material). Each model has been simplified with a voxel resolution of 16^3 and 64^3, and normal maps of 1024^2 pixels were generated. To compare, we used four available applications. The commercial applications *Autodesk Maya 2014* and *Autodesk Mudbox 2014* have different simplification implementations. Furthermore, we made use of *MeshLab*[Cignoni2008] which implements the quadric edge collapse decimation[Garland1997], and we have used *OpenFlipper*[Mobius10] which is based on the general framework for mesh decimation described by Kobbelt[Kobbelt1998]. The simplifications were compared (1) when aiming at equal numbers of triangles and (2) when aiming at similar storage sizes.

4.2 Similar Numbers of Triangles

For rendering purposes, a low number of triangles is desirable. Hence, we first compare the simplification approaches when reducing to a number of triangles similar to our proposed approach (Table 1).

Table 1. Triangle counts. Number of triangles for the proposed solution, given for two resolutions

	CoinPile	JewelPile	Vicky	Totem
Orig.	540,768	77,316	109,905	7,648
64^3	21,596	21,634	4,448	3,046
16^3	1,090	986	454	388

Figure 2 compares several approaches and clearly shows the benefits of our approach. Despite obvious geometric distortions, we reproduce a model which remains recognizable, and visual features are preserved without any manual intervention. One can argue that we need texture data to achieve these results, and the storage size required for these textures is relatively large compared to the simplified mesh. Hence, to have a fair comparison from a storage point-of-view, we have set up a second test aiming at achieving similar storage space usage.

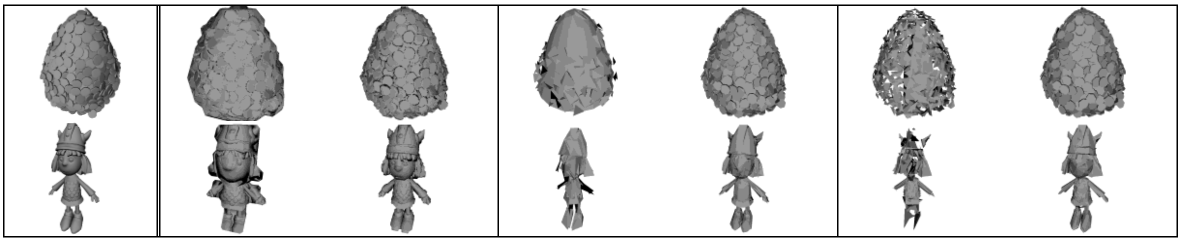


Figure 2. Similar numbers of triangles. Left-to-right: original, proposed (16^3 and 64^3), Meshlab, OpenFlipper

4.3 Similar Storage Sizes

To compare storage sizes, we assume an indexed triangle set. With q indicating the number of bits to represent each coordinate, n_v the number of vertices and n_t the number of triangles, this results in a size of:

$$3\,q\,n_v + 3n_t\,\mathrm{ceil}[\log_2(n_v)] \quad (11)$$

Note that *our* solution requires normal maps and therefore UV-coordinates. These map each vertex to a point in the two-dimensional texture space and require 2q' additional bits per vertex, resulting in a size of:

$$(3q + 2q')\,n_v + 3n_t\,\mathrm{ceil}[\log_2(n_v)] \quad (12)$$

We assume coordinates to be represented using q=q'=32bit floating point numbers. However, our voxelization induces quantization: after 64^3 voxelization, q=6bits suffice to represent the spatial position and after 16^3, q=4bits suffice. After substitution, the numbers of bits required when using 64^3 and 16^3 voxelization are given by Eq. 13 and 14 respectively, resulting in Table 2. In comparison, the number of bits required by the existing simplification approaches is given by Eq. 15, by substituting q=32 in Eq. 11.

$$(18+64)\,n_v + 3n_t\,\mathrm{ceil}[\log_2(n_v)] + size_{tex} \quad (13)$$

$$(12+64)\,n_v + 3n_t\,\mathrm{ceil}[\log_2(n_v)] + size_{tex} \quad (14)$$

$$96n_v + 3n_t\,\mathrm{ceil}[\log_2(n_v)] \quad (15)$$

Table 2. Storage sizes for our approach. This table shows geometry (using Eq.13-14) and texture (PNG-file) size (in KiB)

	CoinPile	JewelPile	Vicky	Totem
Orig.	44,586	7,309	10,452	641
64^3 geom	1,750	1,992	335	220
Texture	2,621	2,592	1,733	1,223
Total	**4,371**	**4,584**	**2,068**	**1,443**
16^3 geom	73	63	28	23
Texture	1,144	989	791	532
Total	**1,217**	**1,052**	**819**	**555**

Figure 3 compares the approaches aiming at equal storage sizes. Even considering the large increase in amount of triangles, features are not fully preserved for the *Vicky* model. For the low-resolution *CoinPile* one can argue that the purely geometric simplification produces a better result, albeit using a much larger number of triangles (see Table 3), impacting the rendering performance.

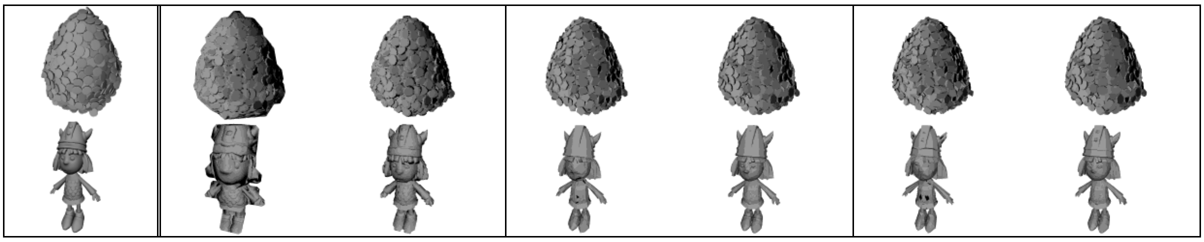


Figure 3. Similar storage sizes. Left-to-right: original, proposed (16^3 and 64^3), Meshlab, OpenFlipper

Table 3. Triangle counts when aiming at similar storage. Given for MeshLab (ML) and OpenFlipper (OF)

	$Pile_{16}$	$Pile_{64}$	$Vicky_{16}$	$Vicky_{64}$
Orig	540,768		109,905	
Prop.	1,090	10,788	454	4,448
ML	8,614	23,863	3,423	7,336
OF	7,212	22,539	4,353	10,432

As final remarks on these results, we note two things. Firstly, as we have analyzed the simplification of just the geometry of an asset, representing the meshes obtained by the suggested simplification approaches does not require the storage of UV-coordinates, which allows a more beneficial storage compared to our approach. However, in general this information will be required nonetheless, resulting in a storage increase of $64n_v$ bits, whereas this information is already present using our approach. Secondly, we note that advanced technologies for texture streaming are already in place to cope with large textural data[Hollemeersch2009]. This allows adaptive fetching of textural data to be done more efficiently than adaptively fetching geometry.

5. CONCLUSIONS AND FUTURE WORK

For simplification of animation film assets we employed an efficient pipeline that generates low-complexity meshes, and we have shown that enhancing the results with normal maps reconstructs the visual appearance. From a rendering performance point-of-view, other simplification approaches are not able to reproduce the same visual quality as our textured models; either they cannot reduce up to the same number of triangles, or they result in heavily degraded models. From a storage perspective, we see that our reduced geometry coupled with a normal map still results in more pleasing results than filling the same storage space with only geometric information. These observations are clearly visible using low-resolution voxelization; at higher resolutions, the better quality as such is debatable, and will depend on the use-case in which the models are employed. Nonetheless we conclude that this approach is a valid alternative for *extreme* asset simplification of static models.

Future work will investigate displacement mapping to further increase the quality, specifically addressing silhouette artifacts. Furthermore, properties other than geometry will be considered, including color and textures. We will also look at how transparency and animations must be taken into account. Transparency precludes the assumption that all parts inside the volumetric representation are permanently hidden. Skeletal animation will pose issues for rigging and will have an influence on the visual quality of the maps which are now generated for *static* models. Finally, if an asset is built of multiple parts for animation purposes, and these animations have to be preserved for interactive environments, these parts cannot be merged.

ACKNOWLEDGEMENT

The research activities as described in this paper were funded by Ghent University, iMinds, the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWOFlanders), and the European Union.

The authors would like to thank Studio 100 NV for the use of their animation models.

REFERENCES

[Akenine2008] Akenine-Möller, T., Haines, E. and Hoffman, N., 2008. *Real-Time Rendering, Third Edition*. A K Peters/CRC Press, Massachusetts, US.

[Alliez2008] Alliez, P., Fabri, A. and Fogel, E., 2008. Computational Geometry Algorithms Library. *Proceedings of the tutorials of SIGGRAPH 2008*. Los Angeles, US, pp. 194.

[AloopingIcon2009] ALoopingIcon. How to remove internal faces with MeshLab. http://meshlabstuff.blogspot.be/2009/04/how-to-removeinternal-faces-with.html. Apr. 2009.

[Autodesk2011] Autodesk, 2011. *The Generation and Display of Normal Maps in 3ds Max* (whitepaper). Autodesk Inc., San Rafael, CA, US.

[Bayever2013] Bayever, J., 2013. *Making of Life Of Pi*. http://www.youtube.com/watch?v=Z-q_MxXf5s.

[Blender2014] Blender 3D: Noob to Pro, 2014. *Creating Pixar-looking eyes in Blender*. http://en.wikibooks.org/wiki/Blender_3D:_Noob_to_Pro/Creating_Pixar-looking_eyes_in_Blender.

[Butts2013] Butts, S., 2013. *Disney INFINITY Goes Mobile*. http://ca.ign.com/articles/2013/08/14/disney-infinity-goes-mobile.

[Cacciola2014] Cacciola, F., 2014. *CGAL 4.5 – Triangulated Surface Mesh Simplification*. http://doc.cgal.org/latest/Surface_mesh_simplification.

[Cignoni1998] Cignoni, P., Montani, C. and Scopigno, R., 1998. A comparison of mesh simplification algorithms. In *Computers & Graphics*, Vol. 22, No. 1, pp. 37–54.

[Cignoni2008] Cignoni, P. et al., 2008. MeshLab: an Open-Source Mesh Processing Tool. *Proceedings of Eurographics Italian Chapter Conference 2008*. Salerno, IT, pp 129–136.

[Coll2010] Coll, N. and Paradina, T., 2010. Accurate Simplification of Multi-Chart Textured Models. In *Computer Graphics Forum*, Vol. 29, No. 6, pp. 1842–1853.

[Disney2014] Disney, 2014. *Disney INFINITY*. https://infinity.disney.com.

[Garland1997] Garland, M. and Heckbert, P., 1997. Surface Simplification Using Quadric Error Metrics. *Proceedings of SIGGRAPH 1997*. Los Angeles, US, pp. 209–216.

[Hazewinkel2001] Hazewinkel, M., 2001. *Encyclopedia of Mathematics*. Springer, Berlin, DE.

[Hollemeersch2009] Hollemeersch, C. et al., 2009. Accelerating Virtual Texturing Using CUDA. *Proceedings of the GPU Technology Conference (abstracts)*. San José, US.

[Khalil2014] El Sayeh Khalil, J. et al., 2014. Transforming Complex Animation Film Assets to 3D Game Assets Through the Use of a Voxel Model. *Proceedings of SCCG 2014*. Smolenice, SK, pp. 24–27.

[Kobbelt1998] Kobbelt, L., Campagna, S. and Seidel, H., 1998. A General Framework for Mesh Decimation. *Proceedings of Graphics Interface 1998*. Vancouver, CA, pp. 43–50.

[Lee2012] Lee, H., Lavoué. G. and Dupont F., 2012. Rate-distortion optimization for progressive compression of 3D mesh with color attributes. In *The Visual Computer: International Journal of Computer Graphics*, Vol. 28, No.2, pp. 137–153.

[Lorensen1987] Lorensen, W. and Cline, H., 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In *Computer Graphics*, Vol. 21, No. 4, pp. 163–169.

[Marras2010] Marras, S. et al., 2010. Controlled and Adaptive Mesh Zippering. *Proceedings of GRAPP 2010*. Angers, FR, pp. 104–109.

[Melax1998] Melax, S., 1998. A Simple, Fast, and Effective Polygon Reduction Algorithm. In: *Game Developer Magazine*, pp. 44–49.

[Mobius2010] Möbius, J. and Kobbelt, L., 2010. OpenFlipper: An Open Source Geometry Processing and Rendering Framework. *Proceedings of the 7th international conference on Curves and Surfaces*. Avignon, FR, pp. 488–500.

[Naylor1990] Naylor, B., Amanatides, J. and Thibault, W., 1990. Merging BSP Trees Yields Polyhedral Set Operations. *Proceedings of SIGGRAPH 1990*. Dallas, US, pp. 115–124.

[Schroeder1992] Schroeder, W., Zarge, J. and Lorensen, W., 1992. Decimation of triangle meshes. *Proceedings of SIGGRAPH 1992*. Chicago, US, pp. 65–70.

[Somers2015] Somers, J., 2015. *Mesh Simplification Viewer*. http://www.jsomers.com/vipm_demo/meshsimp.html.

[Talton2004] Talton, J., 2004. A Short Survey of Mesh Simplification Algorithms. *Course Notes for CS 598 MJG, 2004*.

[Turk1992] Turk, G., 1992. Re-tiling polygonal surfaces. *Proceedings of SIGGRAPH 1992*. Chicago, US, pp. 55–64.

[Turk1994] Turk, G. and Levoy, M., 1994. Zippered Polygon Meshes from Range Images. *Proceedings of SIGGRAPH 1994*. Orlando, US, pp. 311–318.

[Valette2009] Valette, S., Chaine, R. and Prost, R., 2009. Progressive Lossless Mesh Compression Via Incremental Parametric Refinement. In *Computer Graphics Forum - Proceedings of SGP 2009*, Vol. 28, No. 5, pp. 1301–1310.

INFORMATION-ENTROPY BASED LOAD BALANCING IN PARALLEL ADAPTIVE VOLUME RENDERING

Huawei Wang, Zhiwei Ai and Yi Cao
Institute of Applied Physics and Computational Mathematics
No. 2 Fenghao East Road, Haidian District, Beijing 100094, China

ABSTRACT

Aiming at TB-scale time-varying scientific datasets, this paper presents a novel static load balancing scheme based on information entropy to enhance the efficiency of the parallel adaptive volume rendering algorithm. An information-theory model is proposed firstly, and then the information entropy is calculated for each data patch, which is taken as a pre-estimation of the computational amount of ray sampling. According to their computational amounts, the data patches are distributed to the processing cores balancedly, and accordingly load imbalance in parallel rendering is decreased. Compared with the existing methods such as random assignment and ray estimation, the proposed entropy-based load balancing scheme can achieve a rendering speedup ratio of 1.23~2.84. It is the best choice in interactive volume rendering due to its speedup performance and view independence.

KEYWORDS

Load Balancing, Information Entropy, Parallel Volume Rendering, View Independence.

1. INTRODUCTION

With the development of supercomputers and computational simulation, thousands of CPU cores are needed to perform three-dimensional (3D) highly-refined simulations in applied science fields, e.g. inertial confinement fusion and supernova explosion, so as to obtain simulation results of high scientific confidence. The ceaseless increase of computing scale consequentially results in the quick increase of output data amount: the output data usually have several gigabytes at one time step while the total time-varying data may reach the terabyte scale. Moreover, the variables in these data fields may vary violently and behave very differently in different sub-regions. It is a great challenge to quickly and clearly visualize such massive scientific data so as to enable scientists to observe and analyze scientific phenomena effectively.

Volume rendering is one of the kernel algorithms for visualizing 3D data fields (Levoy, 1988, 1990; Drebin et al, 1988; Kaufman and Mueller, 2005). It can directly and effectively exhibit inner data features in high quality. Aiming at large-scale data fields, the parallelization of volume rendering is an inevitable choice: on the one hand, the supercomputer can be used to access massive data; on the other hand, multiple processing cores can be used to share in volume rendering computation and thus accelerate the rendering process. However, in current parallel volume rendering algorithms, especially in adaptive rendering algorithms, the computational imbalance over processing cores results in a great deal of idlesse and thus wastes many computational resources, so the efficiency and the scalability of parallel rendering algorithms are decreased greatly.

In previous work, several adaptive sampling methods were proposed, such as empty space jumping (Drebin et al, 1988), hierarchical adaptive sampling (Ma, 1999; Kähler, 2005), detail-directed sampling (Frisken et al, 2000) and gradient-based sampling (Tang et al, 2008). Kraus et al. presented a GPU-based volume rendering algorithm, where the sampling distance is adaptively chosen according to an oracle (Kraus et al, 2007). In order to grasp variable variation exactly, Marchesin and Verdière proposed a GPU-based cell-projection volume rendering algorithm with adaptive sampling (Marchesin and Verdière, 2009). To deal with TB-scale data output in numerical simulations, it is indispensable to develop parallel rendering algorithms in distributed-memory environments (Ma et al, 1994), but it is difficult to integrate adaptive sampling into parallel rendering due to arbitrary sampling positions. Recently, Wang et al. improved the

ISBN: 978-989-8533-38-8 © 2015

adaptive sampling method in (Marchesin and Verdière, 2009) and proposed a distributed-memory parallel adaptive volume rendering algorithm (Wang et al, 2012). They later introduced information entropy to detect the distribution of physical features and designed a patch-wise adaptive sampling method (Wang et al, 2013).

The uniform sampling method is still used widely in main parallel rendering systems including VisIt (Childs et al, 2006). Though the algorithms in (Ma, 1999) and (Kähler, 2005) are parallel, they use a very simple adaptive scheme depending on only the resolutions of AMR subgrids. In parallel rendering systems, there are multiple factors to influence the rendering efficiency, such as I/O, data structure, load balance, etc. Childs et al. presented a contract-based mechanism allowing many optimizations about these factors to be applied to each pipeline execution (Childs et al, 2005). Moloney et al. proposed an effective dynamic load balancing strategy by pre-estimating the rendering cost of each pixel (Moloney et al, 2007).

In order to improve the rendering efficiency of large-scale data, we propose a static load balancing scheme based on information entropy. In the stage of data assignment, we use the information entropy to estimate the computational amount in ray sampling, and then assign the data patches to the processing cores balancedly according to their computational amounts. All cores will receive a same number of data patches and their total computational amount is nearly equal, which guarantee load balance of the rendering algorithm, thus the rendering efficiency is enhanced and the rendering time is shortened.

2. PARALLEL ADAPTIVE VOLUME RENDERING ALGORITHM

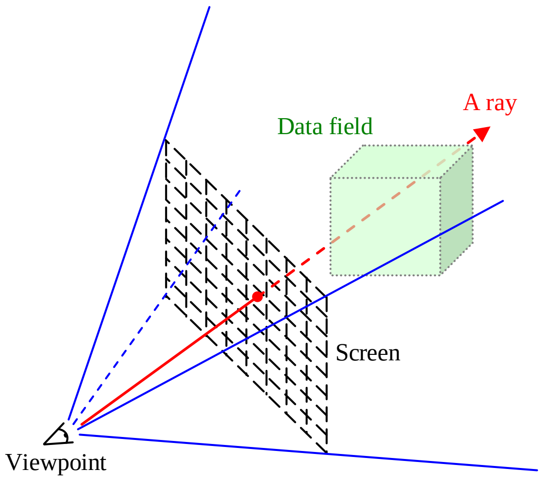


Figure 1. Ray-casting Volume Rendering

A simple description of ray-casting volume rendering is shown in Figure 1. A ray is casted from the viewpoint to each pixel in the screen, along which the data field is resampled. All sampling points will be mapped to color and opacity values from transfer functions. Composing all sample points on each ray according to the depth order, we then obtain the color for the pixel corresponding to the ray. Here, we consider the following parallel adaptive volume rendering algorithm (Wang et al, 2012) (see Figure 2):

1) In a parallel computing environment, the volume rendering engine divides original input data into many small patches and distributes them to all processing cores averagely.

2) On each processing core, the volume rendering engine adaptively samples all distributed data patches on the rays that intersect them: determining positions of sampling points and interpolating values of variables.

3) In the compositing process, the rendering image is partitioned into small blocks and distributed to all processing cores averagely, and each core takes charge of collecting all sampling points and sorting them in depth order for each responsible ray.

4) According to transfer functions and lights, each processing core calculates color and opacity values of sampling points for each responsible ray, and then composites them front-to-back relatively to the viewpoint by using the preintegration method.

5) The main engine collects all image blocks and thus obtains the whole volume rendering image.

The adaptive sampling method in the algorithm is to sample each ray at the extreme points of the rendered variable so that we can exactly grasp the variation trend of the variable to avoid artificial cracks. In detail, the continuous data field in a single cell is defined as trilinear interpolation of its eight vertices, and the variable function along each ray passing through the cell is proved to be a polynomial of degree 3 at most. Thus, the function has at most three monotone intervals, which are separated by extreme points. Solving the extreme points and then taking them as sampling points, we can exactly depict the value and trend properties of the rendered variable.

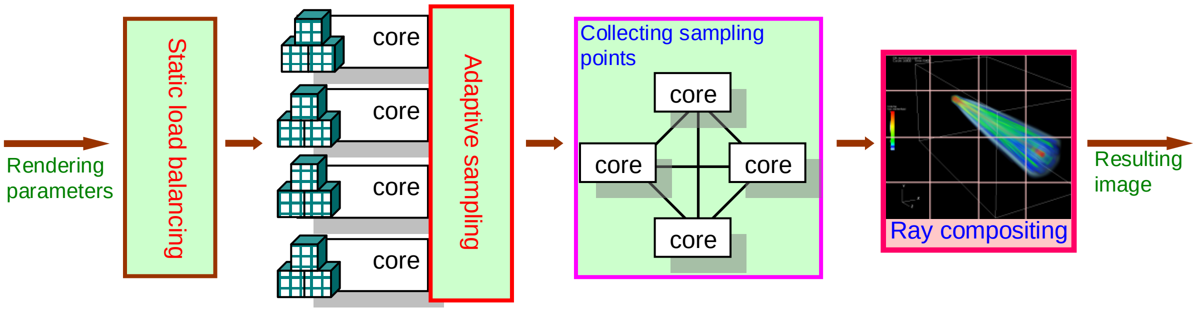


Figure 2. Main Framework of Parallel Adaptive Volume Rendering

3. INFORMATION ENTROPY MODELS

The above sampling method adapts well to uneven distribution of physical features. In the region containing abundant features, the data must change violently and thus need to be sampled densely, while the data can be sampled sparsely in the smooth region containing few features. Hence, we introduce the concept of information entropy (Cover and Thomas, 2006) to depict the abundance degree of physical features, or the degree of violent variation of the data. We then take the information entropy of a data patch as a coarse estimation of the computational amount in ray sampling so as to guide the design of an effective load balancing. In existing work, based on information entropy, Wang et al. identify and present the most essential aspects of time-varying data (Wang et al, 2008), and Wu et al. select important time steps and sub-regions or optimal viewpoints in the visualization process (Wu et al, 2012).

The information entropy is a general measure of information content contained in a dataset, which is defined as

$$H(X) = -\sum_{i=0}^{N-1} p(x_i)\log p(x_i), \qquad (1)$$

where the dataset is regarded as a series of occurrences of a random variable X, which takes values from $\{x_0, x_1, \ldots, x_{N-1}\}$, and $p(x_i)$ denotes the probability of $X = x_i$. In actual applications, a large-scale data field is usually divided into a number of data patches. For each data patch, we can establish the probability distribution function based on a histogram. We partition the variable extent into N intervals, and take the variable falling into an interval as a random event. We then traverse the whole data patch and count the total number of such events for each interval. After normalizing these numbers, we obtain the probability of the variable locating in each interval. From the probability distribution, we then compute the information entropy of the data patch.

Usually there are multiple attributes for a dataset, such as variable value, gradient and derivative. Each of the attributes may be used to depict the information content from certain point of view. Here, we choose two most commonly-used attributes: value and gradient's magnitude, and accordingly we need to construct the probability distribution from a two-dimensional (2D) histogram, as shown in Figure 3. We partition the value's extent into m intervals and the extent of the gradient's magnitude into n intervals and thus obtain an $m \times n$ 2D histogram. Then, we count the probability of each data point falling into the $m \times n$ bins of the histogram, that is to say, its value and gradient's magnitude are all located in the intervals corresponding to a bin. The data points in a data patch can be either zone-centered or node-centered. Using Equation (1), we

ISBN: 978-989-8533-38-8 © 2015

finally compute the information entropy of the data patch from the 2D joint probability distribution. Note that the size of each dimension in the histogram can vary freely and the resulting information-theory models differ slightly. For instance, let $N = m \times n = 256$, then we will test the following cases of $m \times n$ in the experiment: 256×1, 64×4, 16×16, 4×64, 1×256.

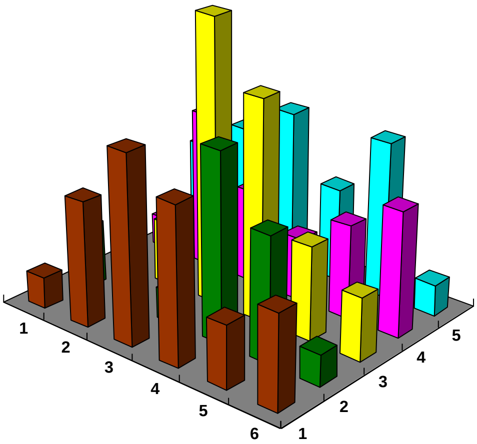


Figure 3. Two-dimensional Histogram

4. ENTROPY-BASED LOAD BALANCING

In parallel rendering, it is very time-consuming to communicate massive data between the computing nodes, thus the static load balancing strategy is usually adopted. The data must be divided and distributed to all nodes at first, and then loaded into the memory of each node. In detail, the divided data patches, which often have the same size, are averagely assigned to all processing cores in a traditional scheme, thus the time of data reading is almost the same between all cores. However, for the parallel adaptive volume rendering, the computing time of ray sampling may differ much for each data patch, and the ray sampling takes a prominent ratio of the rendering time, therefore the load of ray sampling may be very uneven, which will decrease the performance of the parallel rendering algorithm greatly.

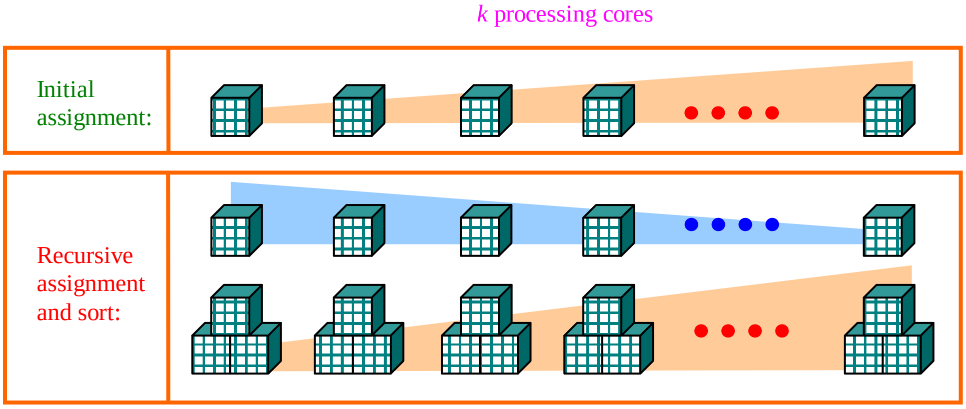


Figure 4. Sketch Map of Data Assignment

To enhance the rendering efficiency, we propose a static load balancing scheme based on information entropy. We take the information entropy introduced in the previous section as the pre-estimated load of ray sampling for each data patch. The information entropy is computed in advance and saved in the dataset as a part of its metadata. When assigning the data patches to the processing cores, the actual data field has not been loaded into the memory, but its metadata has been loaded, thus we can obtain the pre-computed information entropy and use it in load balancing. At first, we sort all the data patches in a descending order according to their pre-estimated loads of ray sampling. Then, we have the load assigning algorithm as shown in Figure 4:

1) Initial assignment. Take the first *k* data patches and assign them to all processing cores respectively in a reverse order, where *k* is the total number of the cores.

2) Recursive assignment of the remainder. Assign the first *k* data patches in the remainder to the cores respectively in turn, i.e. the data patch of bigger load to the core with smaller total load, and then sort the cores in an ascending order according to their total loads.

5. EXPERIMENTAL RESULTS

We experiment with the large-scale dataset output in the 3D laser-plasma interaction hydrodynamic simulation, where the grid size is 512×512×1024, the cell number is 0.27 billion or so and there are totally 150 output time steps. Figure 5 shows the volume rendering image of the data, which clearly reveals the uneven distribution of physical features.

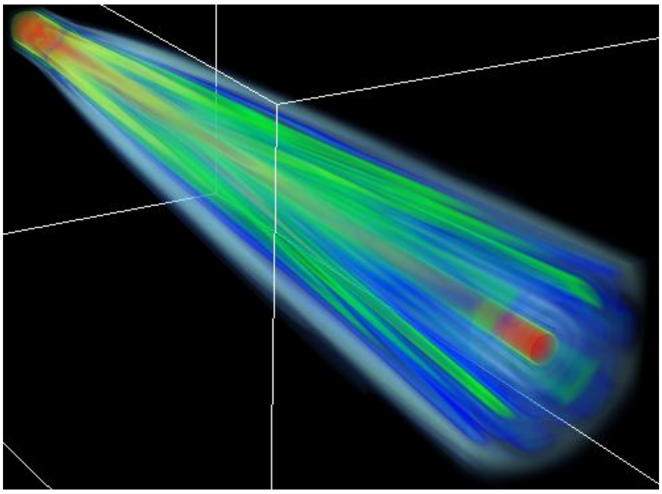


Figure 5. Volume Rendering Image of Laser Filamentation

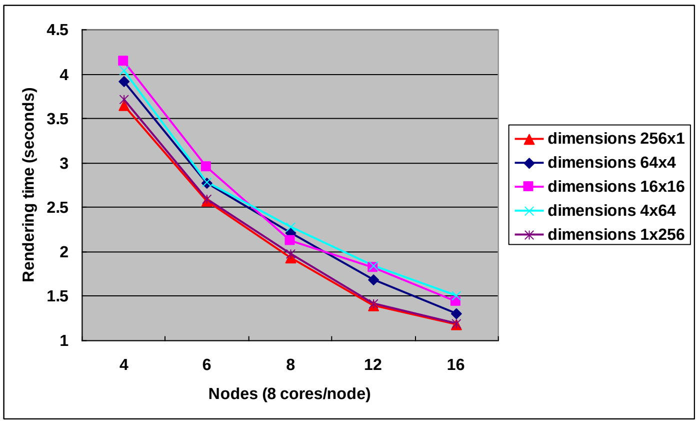
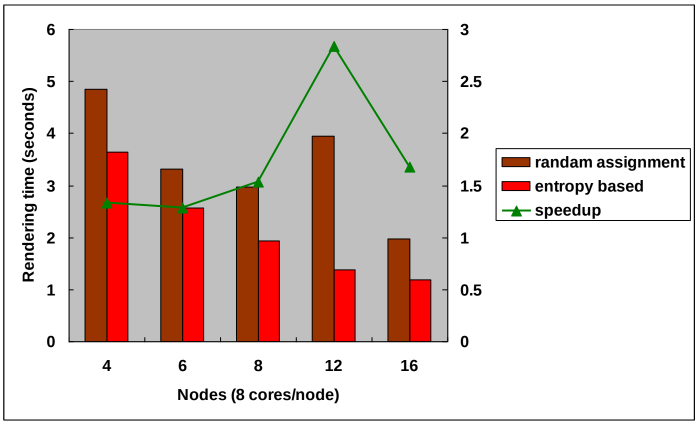


Figure 6. Rendering Time for the Entropy Models of Different Dimensions

We have tested different information-theory models, that is to say, letting two dimensions of the 2D histogram vary from 256×1, 64×4, 16×16, 4×64 to 1×256 and observing the parallel rendering performance at the same time. As shown in Figure 6, the rendering performance behaves similarly with the increase of the parallel scale in these five cases, whose maximal difference is nearly 30%, and the information-theory model based on only value or gradient's magnitude has the best performance.

We also compare the proposed entropy-based load balancing scheme with two existing load balancing schemes: one is random assignment and the other is the ray estimation scheme presented in (Wang et al, 2012), as shown in Figure 7. Relative to the former, our load balancing scheme can achieve a rendering speedup ratio between 1.29~2.84, while the speedup ratio relative to the latter is between 1.23~1.41. Obviously, the parallel rendering performance is very unstable under random assignment. In addition, the ray estimation scheme depends on the viewpoint closely, whose performance will become worse when changing the viewpoint interactively to view the data field. It is clear for static load balancing that the data patches do not move over processing cores in subsequent interaction. Therefore, the proposed entropy-based load balancing scheme not only enhances the rendering speed but also is independent of the viewpoint, thus among these three schemes it is the best choice in interactive volume rendering.



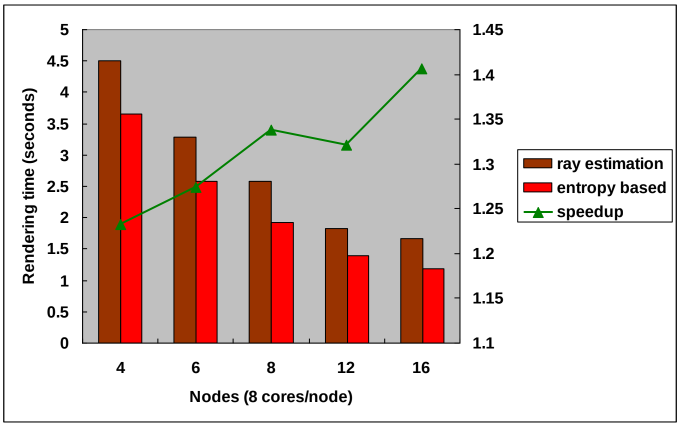


Figure 7. Rendering Times for Different Load Balancing Schemes

6. CONCLUSIONS

This paper presents a novel static load balancing based on information entropy so as to enhance the efficiency of the parallel adaptive volume rendering algorithm. An information-theory model is proposed firstly, and then the information entropy is calculated for each data patch, which is taken as a pre-estimation of the computational amount of ray sampling. According to their computational amounts, all the data patches are distributed to the processing cores balancedly, and thus load imbalance in parallel volume rendering is decreased. Compared with the existing methods such as random assignment and ray estimation, the proposed method can achieve a rendering speedup ratio of 1.23~2.84, and moreover it has the most stable performance in interaction due to its view independence.

ACKNOWLEDGMENTS

The authors wish to thank the anonymous reviewers for their comments. This work is supported by National Natural Science Foundation of China (No. 61232012) and Key Program of Science and Technology Funds of CAEP under grant No. 2014A0403019.

REFERENCES

Childs, H. et al, 2005. A contract based system for large data visualization. *Proceedings of IEEE Visualization '05*, pp. 191-198.

Childs, H. et al, 2006. A scalable, hybrid scheme for volume rendering massive data sets. *Proceedings of EGPGV '06*, Braga, Portugal, pp. 153-162.

Cover, T. and Thomas, J., 2006. *Elements of Information Theory* (2nd Edition). Wiley-Interscience.

Drebin, R. et al, 1988. Volume rendering. *Computer Graphics*, 22(4): 65-74.

Frisken, S. et al, 2000. Adaptively sampled distance fields: a general representation of shape for computer graphics. *Proceedings of SIGGRAPH '00*, pp. 249-254.

Kähler, R., 2005. *Accelerated Volume Rendering on Structured Adaptive Meshes*. PhD Thesis, Freien Universität Berlin.

Kaufman, A. and Mueller, K., 2005. 'Overview of Volume Rendering', in C. Johnson and C. Hansen (eds), *The Visualization Handbook*. Academic Press.

Kraus, M. et al, 2007. Adaptive sampling in three dimensions for volume rendering on GPUs. *Proceedings of the 6th International Asia-Pacific Symposium on Visualization*, pp. 113-120.

Levoy, M., 1988. Display of surfaces from volume data. *IEEE Computer Graphics and Applications*, 8(3): 29-37.

Levoy, M., 1990. Efficient ray tracing of volume data. *ACM Transactions on Graphics*, 9(3): 245-261.

Ma, K., 1999. Parallel rendering of 3D AMR data on the SGI/Cray T3E. *Proceedings of the 7th Symposium on the Frontiers of Massively Parallel Computation*, pp. 138-145.

Ma, K. et al, 1994. Parallel volume rendering using binary-swap compositing. *IEEE Computer Graphics and Applications*, 14(4): 59-68.

Marchesin, S. and de Verdière, G., 2009. High-quality, semi-analytical volume rendering for AMR data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6): 1611-1618.

Moloney, B. et al, 2007. Scalable sort-first parallel direct volume rendering with dynamic load balancing. *Proceedings of EGPGV '07*, Lugano, Switzerland, pp. 45-52.

Tang, W. et al, 2008. An adaptive sampling algorithm based on vector field for medical volume rendering. *Chinese Journal of Shanghai Jiaotong University*, 42(10): 1650-1654. [In Chinese]

Wang, C. et al, 2008. Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6): 1547-1554.

Wang, H. et al, 2012. A parallel adaptive volume rendering algorithm with static load balancing. *Proceedings of IADIS CGVCVIP '2012*, Lisbon, Portugal, pp. 137-141.

Wang, H. et al, 2013. A parallel adaptive volume rendering algorithm based on information entropy. *Proceedings of the 2013 International Conference on Virtual Reality and Visualization*, Xi'an, China, pp. 127-131.

Wu, G. et al, 2012. Information theory in visualization analysis of multivariate time-varying scientific data. *Proceedings of IADIS CGVCVIP '2012*, Lisbon, Portugal, pp. 75-82.

ISBN: 978-989-8533-38-8 © 2015

REAL-TIME VISUALIZING MULTI-FIELD SIMULATION DATA IN EARTH CLIMATE SCIENCE

Yi Cao, Zeyao Mo, Zhiwei Ai, Huawei Wang and Li Xiao
High Performance Computing Center
Institute of Applied Physics and Computational Mathematics
Beijing, China

ABSTRACT

Recent advances in high performance computing technologies are increasing the model complexity, model resolution and the number of simulations in the climate science. Interactive exploration and analysis of these complex, multi-field climate data sets has been identified as one of the major current challenges in scientific visualization today. Without direct 3D multi-field visualization it is hard to see, for example, important correlative effects between vertical wind velocities and transport of volumetric atmosphere. Since the data have complicated 3D structures and are highly time-dependent, the visualization approach must handle this dynamic data in a highly interactive way. In this paper we propose a real-time multi-field visualization framework for earth climate simulation data. A novel visualization pipeline is presented for on-demand data processing, enabling scalable handling large-scale climate dataset. The GPU hardware accelerated multi-field visualization method in the framework allows for the interactive visual analysis of the multiple intersecting climate phenomenons. An information-theoretic based wind field analysis method is also implemented in the visualization framework to help scientist to gain deeper understanding of underlying multi-field climate data.

KEYWORDS

Volume rendering, GPU hardware, multi-field visualization, vector analysis, climate simulation

1. INTRODUCTION

Climate science is a discipline in which scientific progress is critically dependent on the availability of a reliable infrastructure for managing and accessing large and heterogeneous data sets. It is an inherently collaborative and multi-disciplinary effort that requires sophisticated modeling [1] of the physical processes and exchange mechanisms between multiple Earth domains (atmosphere, land, ocean, and sea ice) and comparison and validation of these simulations [2] with observational data from various sources, sometimes collected over long periods of time.

Climate scientists learn from these simulations by comparing modeled and observed data. A variety of grid schemes and temporal and spatial resolutions makes this task challenging even for small data sets. As the need to understand and project climate change becomes increasingly critical, climate model complexity, model resolution and the number of simulations required increases. To keep pace with the growing size and complexity of climate simulations, new software or methods are required to enhance and speed manipulation and analysis of the climate data and model output.

Multiphysics [3] refers to simulations involving multiple physical models such as magneto hydrodynamics, fluid structure interaction, or fluid flow combined with chemical reactions. A coupled climate model is a typical example of a multiphysics model, comprising interdependent models that simulate the Earth’s atmosphere, ocean, cryosphere, and biosphere. Multiphysics simulations are a source of multi-field datasets containing an even larger number of components than are obtained from single field simulations.

The visualization community has produced numerous techniques for looking at single scalar, vector, and tensor fields. However, rather than looking at a single field at a time, scientists need to explore the interaction of these fields to gain deeper understanding of underlying processes and relationships. Creating such tools for exploring these multi-fields [4], time-varying data sets then creates one of the significant challenges in visualization and analytics.

Recent advances in computer hardware such as the Graphics Processing Unit (GPU) [5] are enabling the direct implementation of computationally expensive techniques like volume rendering for the analysis of complex output data sets. But the volume rendering method also leads to the main problem, because conventional coupled rendering method can hardly accurately reflect the deep relationship between the physical phenomena or geometry structures for multi-field data sets.

We present a visually accurate multi-field visualization framework for climate simulation to accomplish the following:

(1) A demand driven visualization pipeline for visual analysis of large data sets, to provide a scalable strategy by reducing massive cost of data filtering and intermediate data generation in the pipeline.
(2) GPU hardware accelerated multi-field visualization for coupling atmospheric volume data with other complex three-dimensional structure of climate models, to implement interactive visual analysis.
(3) Information-theoretic based vector field analysis of climate data, to enhance the visualization quality of arrow representation by extracting the interest characteristics from the vector field.

This multi-field climate visualization system improves the evaluation and prediction of cloud and storm behaviors from simulation data. This system has been applied to global climate change evaluations and weather observer in china.

2. BACKGROUND

2.1 Climate Model

Climate model is an expression of mathematical models of ancient, modern and future climate. It uses quantitative methods to simulate the interaction between the atmosphere, oceans, land surface, and ice. Climate models can be used in many ways, ranging from daily weather to dynamic changes in the climate system over time, and even predictions of future climate. For example, climate models have been used to speculate the consequences of increasing greenhouse gases in the atmosphere, and carbon dioxide (CO_2) is the primary greenhouse gas emitted through human activities. These models predict Earth's average temperature will continue to rise, and the temperature of higher latitudes in the northern hemisphere is one of the fastest growing regions. Global model is a description of the climate of earth as a whole, with all the regional differences averaged, and the climate at a given location on earth is the regional model. The Earth System Grid Federation (ESGF) [6] is an international collaboration for the software that powers most global climate change research, notably assessments by the Intergovernmental Panel on Climate Change (IPCC).

For the climate data visualization, two climate models are introduced in this paper. The Grid-point Atmospheric Model of IAP LASG (GAMIL) [7] and Advanced Regional Eta-coordinate Model (AREM) [8] is the climate model at global and regional scales respectively, and developed by the National Key Laboratory for Numerical Modeling of Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics (IAP), Chinese Academy of Sciences (CAS). GAMIL can be used to describe atmosphere general circulation model based on the finite difference dynamical core. It has been widely used for the studies of the 20th century climate change, seasonal prediction etc, and has taken part in various international model intercomparison projects, such as Atmospheric Model Intercomparison Project (AMIP), Climate of the 20th Century (C20C), Climate Prediction and its Societal Application (CliPas) etc. AREM is designed for high resolution simulating the moisture advection based on regional numerical prediction model. It can handle the actual steep topography in regional climate model. The AREM model has been widely used for the heavy rain forecast in the flood season in China's scientific and business fields, such as meteorology, hydrology and environment.

2.2 Climate Simulation

To improve the numerical fidelity of climate prediction, climate science needs continuously increase the resolution both in modeling and simulation. So the main challenge to realistic climate simulations has been the relatively coarse resolution used in climate models. Due to the extreme computational demands of climate simulations, the average resolutions of global models in IPCC 5th report typically reach 110 km, and the average spatial resolutions of regional models is between 5 km and 10 km.

Fine resolution in climate simulation usually requires high performance computing to handle the complex theoretical models. The high performance GAMIL and AREM are done with J Adaptive Structured Meshes applications Infrastructure (JASMIN) [9], a parallel software infrastructure for scientific computing developed by the Institute of Applied Physics and Computational Mathematics (IAPCM). The main objective of JASMIN is to accelerate the development of parallel programs for large scale simulations of complex applications on parallel computers. With the help of multi-physics parallel computing provided by JASMIN, GAMIL and AREM can complete the simulation on (tens of) thousands of processing cores, and reach the high resolution at 20 km and 8 km scales respectively.

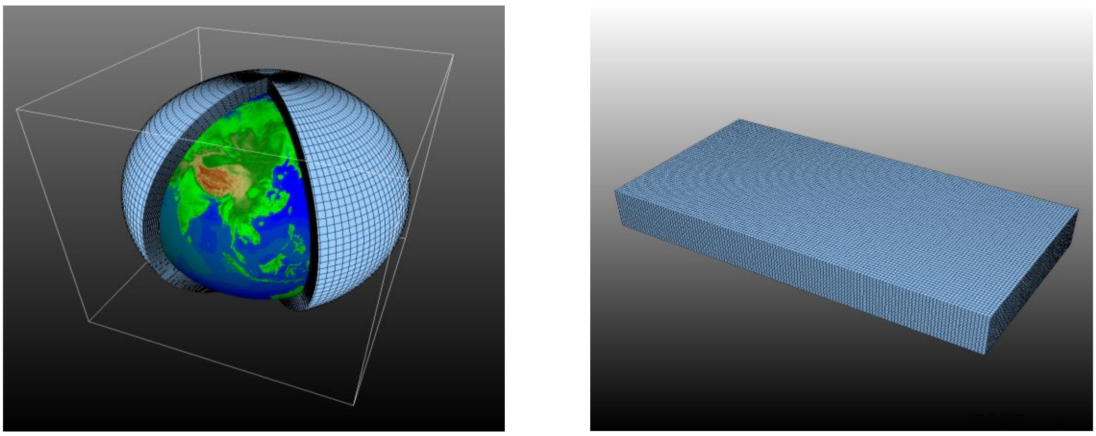
2.3 Climate Data and Visualization

Due to the increase in computing power, climate simulations result in increasingly larger data sets. The data is usually three-dimensional, time dependent and multi-field. As climate models become more complex and output data sets become larger, the raw data can be hard to understand, particularly when the size of the data sets exceeds gigabytes or hundreds of gigabytes. The large-scale climate dataset bring the challenge to interactive visualization and will require re-designing of visualization pipeline. Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT) [10] is an advanced tool well suited for large-scale and complex climate-data analysis problems. On the other hand, climate visualization is still a difficult problem because it comprises multi-field data visualization. The climate models integrated with several equations based on differential laws, such as physics, fluid motion, and chemistry, will output the multi-field dataset. Typical examples are the complex combinations of scalar, vector, tensor fields or also abstract data given over a two to four dimensional space time. Many of these complexities can be subsumed under a concept of multi-fields. So new visualization method need to be developed that can be used to better describe climate data.

3. REAL-TIME VISUALIZATION FRAMEWORK FOR CLIMATE DATA

3.1 Data Description

The climate dataset used in this study was generated from GAMIL and AREM models under JASMIN parallel software infrastructure. GAMIL is modeled based on latitude-longitude grid, and its horizontal grids consist of a uniform zonal grid, and the vertical grids represent the height to the earth ground, as show in Figure 1(a). AREM model is built with E-grid [11], which is used to guarantee high precision calculation of the divergence and vortices. All the data sets from GAMIL and AREM models have been generated in the HDF5 format. In order to help researchers to be more intuitively navigate global climate data, a coordinate transformation from spherical coordinates to Cartesian is executed in this work, as show in Figure 1(b).



(a) 3D climate data in spherical coordinates (b) 3D climate data in Cartesian coordinates

Figure 1. The latitude-longitude grid in GAMIL model

Table 1. The climate dataset used in the experiments

Climate model	Grid resolution	Data size in single time-step
GAMIL	1440×616×26	880 MB
AREM	801×1001×42	900 MB

3.2 The Visualization Pipeline for Visual Analysis of Large Data Sets

High resolution climate simulation will bring massive data sets, which presented a huge challenge to visualization for data analyzing and processing. In order to improve the performance of visualization, we design a demand driven novel visualization pipeline. Similar to VisIt [12], our pipeline employs a modified data flow network. The pipeline components in turn include file reader, data filter and image renderer. Pipeline execution first starts with an update request progress from the data filters to the file reader, here we call it upstream. The next phase in the pipeline is the execution from file reader down to renderer, which we call the downstream.

A dual-granularity strategy of data filtering is used in the visualization pipeline. In the upstream, the block data based pre-filtering strategy is applied. The block data that actually required for reading from disk is only depending on the operation attributes of data filters. All the blocked data needed to be read was recorded the block number of it in the upstream. So we realized the block based on-demand reading. It is mainly through I/O optimization to improve pipeline performance. In the upstream, a cell based fine-filtering strategy is also applied. Each cell in the blocked data will be marked according to the attributes of pipeline filters. That helps the actual data processing for each cell in the downstream. The demand driven mode and dual-granularity strategy enables on-demand data reading in the visualization pipeline, thus greatly reducing the massive cost of filtering data, then saving the system memory resources.

The dual-stream execution can also help to reduce the large amount of intermediate data generation in pipeline. During data filtering, the spatial operation, such as clip, slice and box, will produce a whole unstructured mesh data set compared to original data structure. If the original grid is based on regular or rectilinear grid type, the amount of pipeline data after spatial operation will dramatically increase. In the upstream of visualization pipeline, the mesh cell not involving spatial operation will be marked. So in downstream phase, those marked cells will be unchanged. Due to the small number of units involved the global space operations, so the actual unstructured mesh generated in the pipeline will also relatively small.

3.3 Hardware Accelerated Multi-Filed Visualization

In this paper, the GPU hardware accelerated multi-field visualization is implemented with the deferred shading [13] based multi-filed data fusion framework. Deferred shading is a screen-space shading technique. It is called deferred because no shading is actually performed in the first pass of the vertex and pixel shaders. Instead shading is "deferred" until a second pass. On the first pass of a deferred shader, only data that is required for shading computation is gathered. Positions, normals for each surface are rendered into the geometry buffer (G-buffer) as a series of textures. After this, a pixel shader computes the direct and indirect lighting at each pixel using the information of the texture buffers, in screen space. Deferred rendering can improve the efficiency of the graphics pipeline, because this technology can avoid the high cost of rendering calculations in the final three dimensional objects that are occlued or not visible. The deferred rendering schematic is shown in Figure 2.

In our multi-field visualization, deferred shading is used to control the execution of visualization pipeline. So that the image based multi-filed data fusion is delayed until the ray accumulation stage of GPU volume rendering. That ensures the correct depth sequences and occlusion results in the final image. In the multi-field data fusion framework, each single field is mapped into a plot, each of which is responsible for the rendering of one physical phenomenon. During rendering stage, each plot will be fed into the pipeline in sequence. In the first rendering stage, the framework makes it a priority to send geometry rendering plots into the pipeline. The multi-filed data fusions between them are automatically done by the graphics pipeline based on depth information. In the end of first stage, a serial of intermediate images will be obtained.

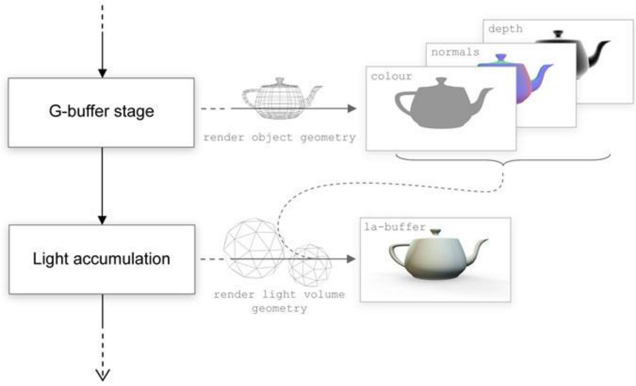


Figure 2. The deferred rendering schematic

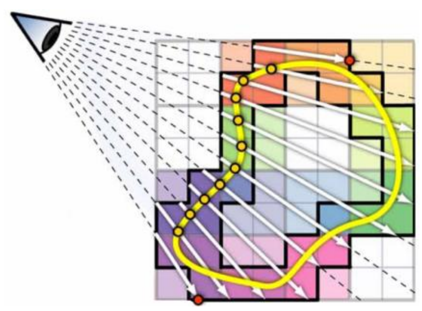


Figure 3. The GPU ray accumulation during multi-filed data fusion

The intermediate images include the RGBA image of geometry and the corresponding depth image. In the second stage, the volume rendering plot, together with the intermediate results generated from the first stage, is send into the pipeline. In order to facilitate the implementation of GPU, all the intermediate images are send in the form of 2D texture. During ray accumulation stage, the two intermediate results will participate in the rendering of final image. In order to realize deferred rendering, the graphics card needed to support some OpenGL extensions, such as Multiple Render Targets (MRT) and floating point textures. Figure 3 depicts the ray accumulation stage during multi-filed data fusion. The rays continuously compare the depth value between current sample point's depth along the ray and the input depth texture. When value of depth texture is greater than the depth of current ray position, that means current ray was obscured by the intermediate geometry. Then the ray can be terminated and perform the blending between volume image and intermediate geometry image.

```
float ConvertLinearDepth( float farPlane, float nearPlane, float d )
{
    return (2*nearPlane)/(farPlane + nearPlane - d*(farPlane-nearPlane));
}
```

Figure 4. The linear depth mapping

In our multi-filed data fusion framework, the depth buffer is very important for it can be used for deferred shading. Using the standard depth, i.e. the value that is obtained by multiplying with World View Projection matrices followed by perspective divide, is often not enough. The main problem with standard depth is that the depth value is not linear. That means 90% of the precision is used up in the first 10 percent of the viewing distance. When the visualization algorithm relies heavily on depth comparison, such GPU based volume rendering, depth artifacts will be caused. That because GPUs rasterize primitives in screen space and interpolate attribute data linearly in screen space as well. The key to handle this problem is to make the depth value linear. However, the non-linear depth cannot be correctly interpolated by simple linear interpolators. According to the near clipping plane (zNear) and far clipping plane (zFar) parameters in perspective projection, we complement the correct linear depth mapping, as can be seen in Figure 4.

3.4 Vector Field Analysis in Climate Data

Vector field is an important component of multi-field climate data. It usually comes from climate dynamics model or atmospheric circulation patterns. Wind field distribution plays a critical role in characterizing a region's weather or the effect of atmospheric circulation on the surface of the Earth. Wind field is also a kind of vector field. To give domain scientists a good insight into the qualitative nature of the vector field, we need to analyze the characteristics of vector field. Topology-based method is a popular method for representation of vector field [14]. Vector field generally have a complex topology. Typical topology structure of vector fields are formed by critical points, as show in Figure 5.

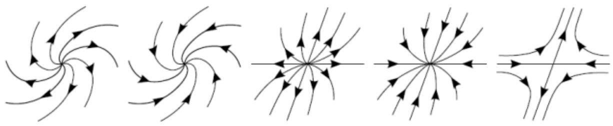


Figure 5. The typical topology structures of vector fields

Arrows representation is a direct visualization method, which is commonly used to reflect the structure of vector field. However, the arrows method also has its flaws. The arrows will lead to icon clustering, causing confusion of the image. Therefore, if the interest characteristics or distribution can be extract from the vector field, we can solve the contradiction between the accurate descriptions of vector field and confused images caused by overlapping icons. Based on this idea, we study the feature-driven adaptive arrows visualization methods. This paper present a statistical analysis method based on angular distributions of vector field. Comparing with topology analysis method, our method can address the problem of uneven distribution of characteristics extracted from vector field.

We apply Shannon's entropy [15] to measure the information content of a vector field. Entropy is used to measure the uncertainty of a random variable. Given a random variable X with a sequence of possible outcomes x, x∈ $\{x_1, x_2, \ldots, x_n\}$, then the information content for the random variable can be computed using Shannon's entropy as:

$$H(X) = -\sum_i p(x_i) \log p(x_i) \tag{1}$$

Where *p(x)* is the probability mass function.

If use the vector direction as a random variable *X*, the extreme range of entropy can be described as follows: If all the vector direction is the same, then *H(X)* reaches a minimum, i.e., it contains a minimum amount of information. This represents a vector fields without critical points. If vectors distributed in each direction, and have the same probability distribution, *H(X)* reaches a maximum, which represents the maximum amount of information it contains. In this case, the vectors in regions will contain characteristics of critical points.

To construct the probability distribution function of the vector field, we use the angle of the vector at each grid node as a random variable. $C(x_i)$ is the number of vectors in bin xi. The probability of the vectors in bin x_i can be computed using Equation 2:

$$P(x_i) = \frac{C(x_i)}{\sum_{i=1}^{n} C(x_i)} \tag{2}$$

Similar to the work in [15], we define the probability sample space as a 360 degree circular for two-dimensional vector field. The circular consists of a finite number of fans (bins) x_i, $i = 1...n$. By mapping two dimensional vectors into the bins in circle, we can approximate the probability mass function of 2D vector field. But there's something different in 3D vector field. We define the probability sample space in unit sphere with n equal area surfaces by using plane triangle rather than spherical polygon of arbitrary shape. Our approximate method can reduce the high computational overhead of 3D vector classification. Through mapping 3D vectors into triangle surfaces, we can obtain the probability mass function of 3D vector field.

4. RESULTS AND DISCUSSION

The multi-filed climate data visualization was run on Inspur's high performance server system that suits large-scale high performance computing in science and engineering computing field. The graphics card in the server node is an NVIDIA Quadro K6000 with 12 GB of RAM and 2880 stream processors. The core volume raycasting was written in C++ and GLSL shaders. Our visualization pipeline supports the remote visualization mode, that is, the massive computing of visualization are performed on a remote server node, and users can interactive explore and analyze the results on their local desktops.

Figure 6 shows the global climate result that visualized with the multi-field visualization framework presented in this paper. The figure contains three different data fields such as geographic altitude, distribution of atmospheric water vapor in January and the wind field at 850 hPa.

ISBN: 978-989-8533-38-8 © 2015

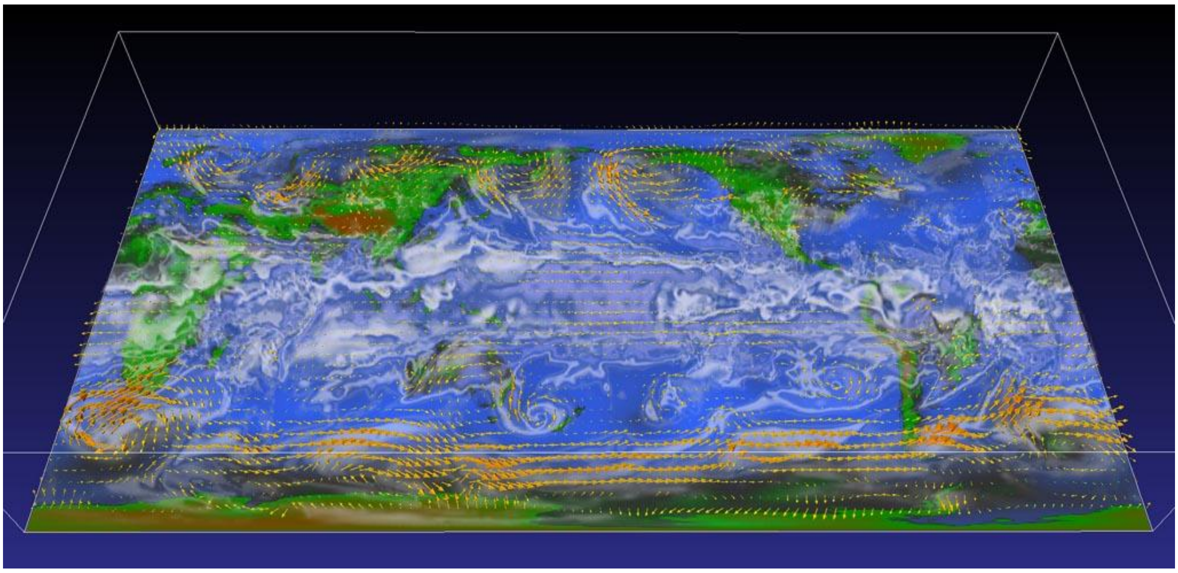
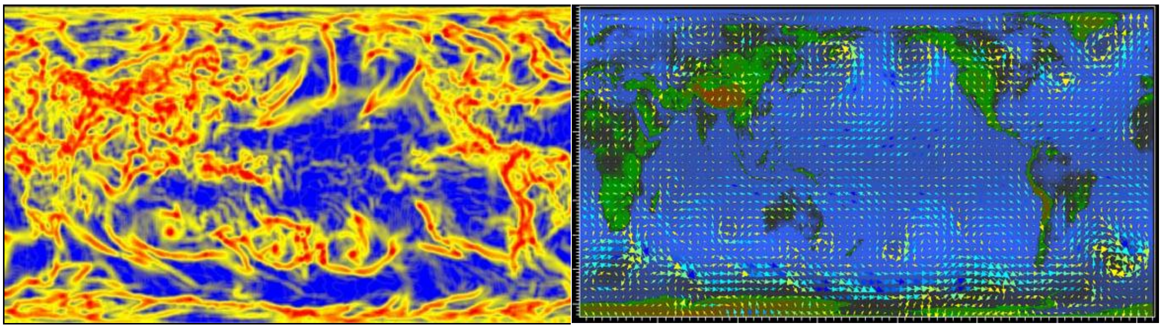


Figure 6. The multi-filed data visualization result in global climate model

The volume rendering method is used to describe the translucent water vapor data. Thanks to the visual means of multi-field coupling, we can clearly observe the correlative effects between the transport of volumetric atmosphere and wind velocities at spatial resolutions of 20 km.

Cloud simulation is always an important and difficult component in climate model. With the help of our visualization framework, the high quality results can clearly exhibit the high resolution cloud simulation ability of GAMIL model. Figure 6 show our visualization method can reasonable represent important weather processes, such as mesoscale eddy activities at high latitude in the southern hemisphere, strong westerlies and storm activities in the Southern Ocean, equatorial easterlies and equatorial clouds.



(a) (b)

Figure 7. The entropy distribution of wind field characteristics in GAMIL model are shown in (a), the arrow based visualization result are shown in (b), the characteristics of vectors is mapped into the arrow colors

Figure 7 (a) show the entropy distribution of wind field characteristics extracted from the GAMIL model data through the information-theoretic based data analysis method in our visualization framework. Observation shows that the entropy will be larger, the red areas in Figure 7 (a), when wind field reflecting the characteristics such as vortex phenomena. In contrast, when the entropy value of wind field is small, as shown the blue areas in Figure 7 (a), that always means the corresponding vortex phenomenon will not be obvious. This corresponding relation between entropy and characteristics of the wind field can be seen by comparing Figure 7 (a) with Figure 7 (b). With the help of entropy-based method for visualization, scientists can be less influenced by the massive and confused vector symbols, and they can be more objective understanding of the global wind distribution patterns in the big simulation data.

Figure 8 show the 3D regional weather result of Typhoon Rananim (2004) that simulated by AREM model and visualized with the multi-field visualization framework presented in this paper. Typhoon Rananim made first landfall in Shitang Town, Wenling City (a sub-city of Taizhou City), Zhejiang Province on August 12. The typhoon was the strongest to hit Zhejiang since Typhoon Sally in 1956.

With our multi-field visualization method, the distribution and movement of the Typhoon cloud systems can be clearly represented. Volume rendering method is used to describe the red translucent Typhoon cloud systems. Figure 8 shows that the position of the vortex center of Typhoon can be predicted by the 48 hours weather forecast generated from high resolution AREM model. Figure 9 (a) show the distribution of vertical thickness of cloud system in near-earth space and the corresponding precipitation intensity distribution on the ground. Figure 9 (b) is the satellite cloud image captured by NASA. Based on the visualization pipeline for on-demand data processing presented in this paper, our visualization framework can implement interactive visual analysis for both global and regional climate data sets.

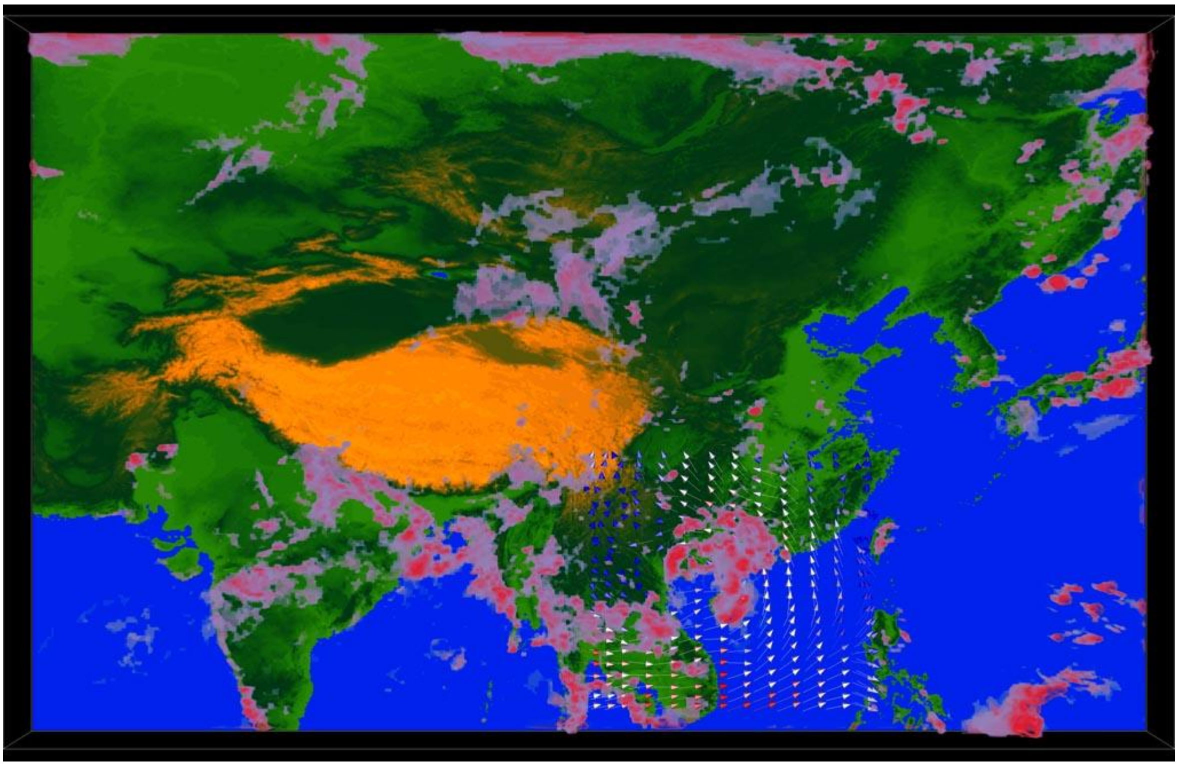
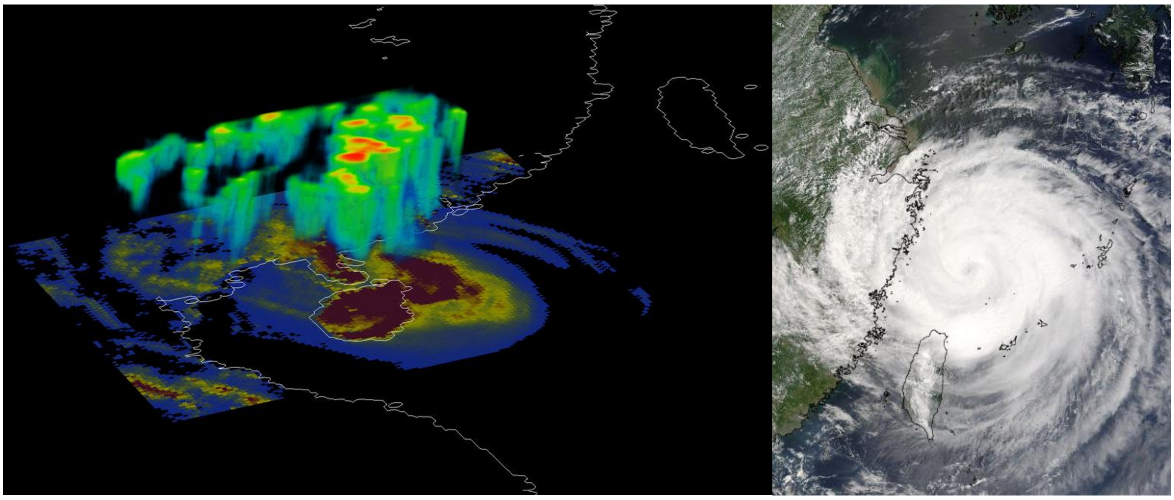


Figure 8. The multi-field weather data visualization of Typhoon Rananim (2004)



(a) (b)

Figure 9. (a) The distribution of vertical thickness of cloud system and precipitation, (b) the satellite cloud image

5. CONCLUSION

Climate scientists learn from simulations by comparing the modeled and observed data. A variety of grid schemes and temporal and spatial resolutions makes the climate visualization task challenging even for small data sets. The multi-field climate data visualization framework presented in this paper has been tested by two applications, GAMIL and AREM. The user case study in those two applications proved the multi-field data fusion method we presented is useful and effective for large data sets. GPU accelerated volume rendering is a key methods to be involved in the multi-field visualization. We need to additionally study the scalable strategies for GPU based multi-field visualization method such as feature-driven data reduction, multi-resolution rendering and parallel rendering.

ACKNOWLEDGEMENT

This work is supported by the Key Program of Science and Technology Funds of CAEP under grant NO.2014A0403019, and the National Natural Science Foundation of China (No. 61232012).

REFERENCES

1. T. Kindler, K. Schwan, D. Silva, M. Trauner, F. Alyea. A parallel spectral model for atmospheric transport processes. Submitted to Concurrency: Practice and Experience, 1994.
2. W.M. Washington and C.L. Parkinson. An introduction to three-dimensional climate modeling. Oxford University Press, 1986.
3. Paul Lethbridge. Multiphysics Analysis, p26, The Industrial Physicist, Dec 2004/Jan 2005.
4. Keherer J., Hauser H. Visualization and Viusal Analysis of Multi-faceted Scientific Data:a Survey.IEEE Transactions on Visualization and computer Graphics, 19(3):495-513,2013.
5. Amburn, P., Berberich, M., Moorhead, R. J., Dyer, J., & Brill, M. (2009). Geospatial Visualization Using Hardware Accelerated Real-Time Volume Rendering. IEEE Oceans 2009.
6. Luca Cinquini, Daniel Crichton, Chris Mattmann,et al, “The Earth System Grid Federation: An Open Infrastructure for Access to Distributed Geospatial Data”, IEEE Future Generation Computer Systems, http://dx.doi.org/10.1016/j.future.2013.07.002, 17 September. 2013.
7. Li, L. J., and Coauthors, 2013a: Evaluation of grid-point atmospheric model of IAP LASG version 2 (GAMIL 2). Adv. Atmos. Sci., 30, 855-867, doi: 101007/s00376-013-2157-5.
8. WU Shuiqing, XU Youping, HU Banghui,et al. The application and experimentation of a new hydrostatic extraction of reference atmosphere in AREM [J].Torrential Rain and Disasters, 2013, 32(2): 132-141
9. Zeyao MO, Aiqing ZHANG, Xiaolin CAO, Qingkai LIU, Xiaowen XU, Hengbin AN, Wenbing PEI, Shaoping ZHU. JASMIN: a parallel software infrastructure for scientific computing. Front. Comput. Sci. China. 4(4): 480-488, 2010.
10. UV-CDAT home page. http://www.uv-cdat.org/
11. Yu Rucong. Applicat ion of a Shape- Prreserving Advection Scheme to the Moisture Equat ion in an E- grid Regional Forecast Model. Adv Atmo S ci, 1995, 12(1) : 13~ 19.
12. VisIt home page. https://wci.llnl.gov/simulation/computer-codes/visit/
13. Hargreaves, S., Harris, M., 2004. Deferred rendering. NVIDIA Corporation. 44 pp.
14. A Pobitzer, R Peikert, R Fuchs, B Schindler, A Kuhn, H Theisel. The State of the Art in Topology-Based Visualization of Unsteady Flow. Computer Graphics Forum 30 (6), 1789-1811, September 2011.
15. Lijie Xu, Teng-Yok Lee, Han-Wei Shen, "An Information-Theoretic Framework for Flow Visualization", IEEE Transactions on Visualization & Computer Graphics, vol.16, no. 6, pp. 1216-1224, November/December 2010, doi:10.1109/TVCG.2010.131.

MULTI-BOUNDARY SHAPE RETRIEVAL BASED ON A NEW CLASS OF MOMENT FUNCTIONS

Ruixia Song[1], Xiaochun Wang[2], Yena Wang[1] and Mei Gu[1]
[1]*College of Sciences, North China University of Technology, Beijing, China, 100144*
[2]*College of Sciences, Beijing Forestry University, Beijing, China,100083*

ABSTRACT

Based on a class of complete orthogonal function system, V-system, this paper proposes a new kind of moment functions (called V-moment functions), and applies them to the shape retrieval. The V-moments are orthogonal, and involve only simple linear operations. The V-moments can be used to extract image features accurately, and the original image can be reconstructed with only a small amount of them. The V-moments have advantage in extracting features of image with complex boundaries since the V-system contains a great deal of discontinuous basis functions. Therefore, feature extraction of multi-boundary image using V-moments is very promising. This paper performs image retrieval based on their shape features. The results of retrieval experiment, conducted on benchmark database MPEG-7-shape-CE2, show that the algorithm proposed in this paper outperforms some classical moments including Zernike moments, Hu invariant moments, orthogonal Fourier-Mellin moments, Legendre moments and the geometric central moments in retrieval efficiency according to several evaluation indexes.

KEYWORDS

Content based image retrieval; Multi-boundary shape retrieval; Orthogonal moment functions; V-system; V-moment.

1. INTRODUCTION

With a dramatic increase in the number of image in the internet and the spread of computer, image retrieval technique has been developed greatly. The development of computer-based image retrieval experienced the process from text annotation based method to image content based one. Content-based image retrieval is the process of retrieving the related images, those similar to the query, from large database collections based upon specific visual characteristics. The key technique of image retrieval is image feature description, which directly affects the image similarity judgment.

Image features generally include color, texture, shape, etc. In this paper, we focus on the research of image shape feature. In general, the shape feature description can be divided into two categories, boundary-based and region-based. The boundary-based description techniques are often applied on the shapes whose boundary information is easy to get, the widely used methods include Fourier descriptor, curvature scale space descriptor, wavelet descriptor and shape context, etc. When using the boundary-based description techniques, the computing quantity can be reduced and the processing becomes more convenient since we only need to process the boundary information of a shape. It should be noted that when the boundary of a shape is not a whole closed curve but a curve consisting of several closed ones we couldn’t obtain accurate shape features if using the continuous feature descriptors. In addition, boundary extraction result depends on the image segmentation technology, and is sensitive to noise and slight changes on the boundary. Therefore, using only the boundary features to describe the overall characteristics of the whole image is usually not accurate enough. However, when shape is complicated and the whole boundary information is difficult to get, the boundary-based technique is not suitable for shape retrieval, while the region-based approaches, which takes into account the whole image region rather than only use boundary information, are more reliable. Therefore, for the images with complex boundaries (such as some trademarks, some logos, some symbols, etc), we usually use the region-based techniques. Method of moments is a typical region-based feature description method, in which the statistical features of the region pixels, i.e. the whole region features are described. The well-known Hu invariant moments, Zernike moments and Fourier-Mellin moments are widely used in shape retrieval. Among these moment shape descriptors, Zernike moments are the most desirable for shape description.

Moment function has been proven to be a very effective tool to extract the shape feature of images. The earliest application of moment function in image analysis can be found in the literature published in 1962 by Hu [1]. Hu obtained seven moment invariants using geometric moments, and applied them to the feature recognition. Theory of moment function has developed rapidly since 90's of the last century, and many methods have been developed. The well-known Legendre moments, Fourier-Mellin moments and Zernike moments [2]-[5] have wide application in many areas, and greatly promote the development of pattern recognition.

In general, for complex image retrieval, People usually employ the region-based method. In recent years, multi features fusion method attracts more and more attentions. The typical method includes the combination of shape features and topological features [6], the combination of shape descriptors and feature matching [7], the combination of global features and local features [8-9], the combination of three features of shape, color and texture [10], the combination of overall shape and internal structure [11], each of these algorithms is ideal for certain types of images. Many algorithms in single boundary image retrieval achieved good results, but often they are not suitable for retrieval of multi-boundary image. For example, reference [12] uses polygonal approximation and cubic polynomial curve fitting to represent the outline shape and defines shape similarity based on similarity of resulting curves, because it fits the outline of an image with a closed curve, it is not suitable for the multi-boundary image. Recently, reference [13] proposes an affine invariant characteristic ratio (CHAR) for complex symbols, and uses it to describe shape of some complex symbols, obtains satisfactory symbol recognition results. In a word, multi-boundary image retrieval (or complex image retrieval) has been paid more and more attention, it is very difficult and challenging work.

In this paper, a new class of orthogonal moments, V-moments, is proposed based on the orthogonal functions system, V-system [14], and is applied to the multi-boundary image retrieval. Compared with classic orthogonal moments, the V-moments have advantages in multi-boundary image feature extraction, and the V-moments have better reconstruction power for complex image due to the V-system has basis functions with discontinuities. Hence, the V-moments can be used to describe multi-boundary image feature better, i.e. the more accurate feature of complex image can be obtained using V-moments.

2. THE V-SYSTEM

The V-system of degree k is a complete orthogonal function system with *k*th order piecewise polynomial as its basis functions. For more details on the V-system, please refer to Song *et al* [14]. In this paper, we only present the expression of the V-system of degree one.

The V-system is composed of groups and classes. The first group comprises of the first two Legendre polynomials on interval [0, 1]. Their expressions are

$$V_1^1(x)=1,\ \ V_1^2(x)=\sqrt{3}(1-2x);$$

The second group consists of two orthogonal generators, which are the most important two functions because they will generate two classes of the *n*th ($n \ge 3$) group accordingly. They are defined as

$$V_2^1(x)=\begin{cases}\sqrt{3}(1-4x), & 0\le x<1/2,\\ \sqrt{3}(4x-3), & 1/2<x\le 1.\end{cases},\qquad V_2^2(x)=\begin{cases}1-6x, & 0\le x<1/2,\\ 5-6x, & 1/2<x\le 1.\end{cases}$$

The values at knot *x*=1/2 are the arithmetic averages of the left and the right limits of the function at *x*=1/2, so $V_2^2(x)$ is continuous, and $V_2^2(x)$ is discontinuous at *x*=1/2.

The *n*th ($n \ge 3$) group consists of two classes, each class comprises of 2^{n-2} functions obtained by squeezing, translating and duplicating one of generators in the second group. The general expression is

$$V_n^{i,j}(x)=\begin{cases}\sqrt{2^{n-2}}V_2^i[2^{n-2}(x-\frac{j-1}{2^{n-2}})], & x\in(\frac{j-1}{2^{n-2}},\frac{j}{2^{n-2}})\\ 0, & others\end{cases}$$

where $i=1,2, \quad j=1,2,\cdots,2^{n-2}, \quad n=3,4,5,\cdots$. Function system

$$\{V_1^1(x),V_1^2(x),V_2^1(x),V_2^2(x),V_n^{i,j}(x), n=3,4,\cdots; i=1,2; j=1,2,\cdots,2^{n-1}\}$$

is the V-system of degree one, an orthogonal function system consisting of infinite number of piecewise linear functions. Chao *et al* [15] proved that the V-system is a class of multi-wavelet. The V-system with multi-resolution characteristic has advantage in expressing object features. We also notice that since the V-system contains a great deal of discontinuous basis functions, which makes it has more advantage in extracting features of separation group object compared to continuous orthogonal function system.

For the convenience of following discussion, the V-system is arranged as $\{V_0(x),V_1(x),V_2(x),\cdots\}$, where

$$V_0(x)=V_1^1(x), V_1(x)=V_1^2(x), V_2(x)=V_2^1(x),\cdots\cdots .$$

3. THE V-MOMENTS

Geometric moments are typical representatives of the moment functions. Based on the geometric moments, we define the V-moments, which is the key technique used in this paper.

The $(p+q)$th order geometric moments of function $f(x,y)$ on G are given by

$$m_{pq}=\iint_G x^p y^q f(x,y)dxdy, \quad p,q=0,1,2\cdots .$$

The Hu moments [1], which are constructed based on the geometric moments, have been used in many applications. It is noteworthy that the geometric moments are based on the function system $\{x^i, \ i=0,1,2,\cdots\}$. However, this function system has two disadvantages: (1) It is not orthogonal, the geometric moments are accordingly not orthogonal moments. However, orthogonality principle is very important in signal processing, especially in signal feature extraction, because orthogonal moments have minimum information redundancy when it is used to represent signal, and the original signal can be almost accurately reconstructed through the inverse transform, so more accurate features can be extracted. (2) As order of geometric moments increases, the operation complexity increases dramatically, it is unfavorable for operational efficiency.

As the V-system of degree one comprises orthogonal piecewise linear functions, the corresponding V-moments could overcome the above two shortcomings. Hence, we define the V-moment of $f(x,y)$ by

$$\mu_{pq}=\iint_G V_p(x)V_q(y)f(x,y)dxdy, \quad p,q=0,1,2\cdots ,$$

where function $V_i(x)$ is the ith basis function of the V-system in the sequential order. Because the V-system is constructed by groups and is uniformly convergent in group's order, the basis function should be chosen by groups (each group as a unit) when constructing the V- moments. We call the V-moments formed by basis functions in the first mth group the V-moments of order m. In addition to the advantage of "orthogonality" and "low computing complexity", the V-moments have the characteristic of "discontinuity". Compared with the classical continuous orthogonal function system, the V-system is composed of not only smooth functions, such as $V_0(x),V_1(x)$, also a large number of functions with different discontinuities $V_n^{2,j}(x), n=2,3,\cdots,$ $j=1,2,\cdots,2^{n-2}$. This kind of discontinuity is favorable when it is used to represent a complicated image with multiple separate boundaries.

Now we use the V-moments for image expression. For an image with brightness function $f(x,y),\ x=0,1,\cdots,M-1,\ y=0,1,\cdots,N-1$, its V-moments are defined by

$$\mu_{pq}=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}V_p(x)V_q(y)f(x,y)=V_pFV_q^T ,$$

where F is a $M\times N$ matrix formed by the image brightness function, i.e $F=\left(f(x,y)\right)_{M\times N}$, V_p and V_q are two row vectors:

$$V_p=(V_p(0),V_p(\frac{1}{M}),\cdots,V_p(\frac{M-1}{M})), \quad V_q=(V_q(0),V_q(\frac{1}{N}),\cdots,V_q(\frac{N-1}{N})) .$$

For convenience, we set matrix $\Delta=(\mu_{ij})_{m\times n}$, then $\Delta=\bar{V}_M F \bar{V}_N{}^T$, where $\bar{V}_M$ is a $m\times M$ matrix made up of the row vectors $V_0,V_1,\cdots V_{m\text{-}1}$, and $\bar{V}_N^T$ is a $N\times n$ matrix formed by the column vector $V_0^T,V_1^T,\cdots V_{n\text{-}1}^T$. Using matrix multiplication, we can find out all V-moments quickly.

Due to the orthogonality of the V-momets, the original image can be reconstructed by the obtained V-moments. The reconstruction formulary is

$$f(x,y)=\sum_{p=0}^{m-1}\sum_{q=0}^{n-1}\mu_{pq}V_p(x)V_q(y).$$

With the use of matrices, it can be rewritten as $F=\bar{V}_M^T\Delta\bar{V}_N$.

It is well known that Zernike moments and Legendre moments are the most effective for feature extraction in all classical moments techniques. Fig.1 shows reconstruction results of an image (Fig.1(a)) using V-moments (Fig.1(b)-(e)), Zernike moments (Fig.1(f)-(i)) and Legendre moments (Fig.1(j)-(m)) with different orders, where the number after each capital letter V, Z and L is the order of V-moments, Zernike moments and Legendre moments, respectively. Obviously, the lower moments depict the image's outline, and with the order increasing, the reconstructed image becomes more and more accurate. We can see that the V-moments have the best reconstruction effect; the image is almost accurately reconstructed using V-moments of order 5, and only use 16 basis functions.

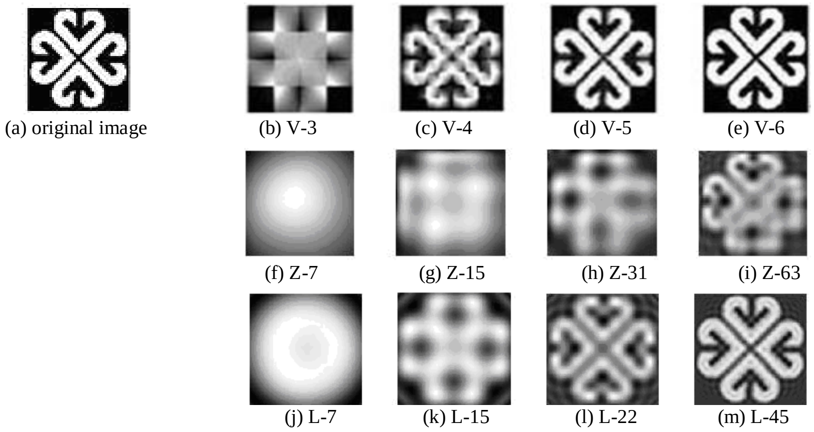


Figure 1. Reconstruction results of an image using V-moments, Zernike moments and Legendre moments with different orders (The numbers after capital letters are the orders of the moments)

4. IMAGE PREPROCESSING

For a given image, it should undergo preprocessing operations before performing image retrieval, since image preprocessing can improve the retrieval result. Image preprocessing commonly comprises a series of sequential operations, including color to gray conversion, image binaryzation, principal direction normalization, and scale normalization. The principal direction normalization includes shifting the centroid of an image to the origin and rotating the spindle to the coordinate axis direction. The scale normalization means normalizing the image to a same size. The following is the process of principal direction normalization and scale normalization.

Assume (x_0, y_0) is the centroid of a given image. According to K-L transform, the direction θ of the eigenvector corresponding to the largest eigenvalue can be calculated as:

$$\theta=\frac{1}{2}\tan^{-1}\frac{2m_{11}}{m_{20}-m_{02}}$$

where $m_{kr} = \sum_{x,y}(x-x_0)^k(y-y_0)^r f(x,y)$ is the centroid moment of order $k+r$. By performing image transform:

$$\begin{cases} x' = x\cos\theta - y\sin\theta, \\ y' = x\sin\theta + y\cos\theta, \end{cases}$$

the spindle of the image is rotated to the x coordinate axis.

Note that even after principal direction normalization, the spindle orientations of two similar images may differ by 180 degrees, and the directions of their vertical spindles, which are perpendicular to the spindles, may differ by 180 degrees too. As shown in Fig. 2, which is composed of four similar images, some have opposite spindle orientations (images exhibit the left-right reversal), some have opposite vertical spindle directions (images are the up-down reversed), and some have both opposite spindle orientations and opposite vertical spindle directions (images exhibit both left-right and up-down reversal). These uncertainties may result in misjudgment in image recognition. To avoid this kind of misjudgment, we flip the principal direction normalized images left to right, bottom to top, and both left to right and bottom to top, and obtain four images with different directions (as shown in Fig. 3). When measuring the similarity between image A and image B, we will calculate the similarities between each of the four images obtained by flipping A and image B, then choose the maximum similar value as the similarity between A and B.

After shifting centroid of the image to the origin and performing image principal direction normalization, we then find its external rectangle by projecting the target image to the two coordinate axes. Keeping the centroid of the target image at the origin, the external rectangle is farther enlarged to a square with length of side being equal to the external rectangle's longer sides, and the square domain is extracted. The last step of preprocessing is standardizing image size so that they have a uniform size 128×128.

After performing preprocessing above, the image becomes better to image translation, scaling and rotation in terms of robustness.

(a) Four similar images

(b) Images obtained by performing principal direction normalization to the images in (a)

Figure 2. The similar images with inconsistent principal direction

(a) (b) (c) (d) (e)

Figure 3. Images obtained by performing up-down and right-left flip.

(a) Original image, (b) Principal direction normalization, (c) Up-down flip, (d) Right-left flip, (e)Up-down and right-left flip

5. IMAGE FEATURE DESCRIPTION USING V-MOMENTS AND SIMILARITY

For a given digital image, we first translate the image centroid to the origin, and choose appropriate value of m and n, then compute the V-moments matrix Δ, next arrange its elements in a zigzag scan way and obtain a vector $\{\mu_{00}, \mu_{10}, \mu_{01}, \cdots, \mu_{mn}\}$, which is a feature vector of the image.

The similarity between two images is measured by the Euclidean distance between the acquired V-moments feature vectors. The smaller distance means high similarity, and image classification and retrieval can be performed based on the distance.

It is necessary to perform preprocessing on the entire database images $\{B_k\}$, and store all V-moments feature vectors $\{\Gamma_{B_k}\}$ in a database before image retrieval. This process can be performed off-line. For a query image A, we just need to compare it with the entire feature vectors in the database. The detailed process is as follows.

Step 1: Input the query image A, and perform image preprocessing including image binaryzation, principal direction normalization and scale normalization described in section 4.

Step 2: Perform right-left, up-down flip to the image obtained in step 1 (as shown in Figure 3), thus there exist four images A1-A4 corresponding to image A.

Step 3: Compute V-moments of images A1-A4, and obtain the feature vector

$$\Gamma_{\mathrm{Ai}} = \{\mu_{00}^{i}, \mu_{10}^{i}, \mu_{01}^{i}, \cdots, \mu_{mn}^{i}\}, \quad i = 1, 2, 3, 4$$

Step 4: Calculate the similarity between images A and any image B_k in the database, which is denoted by $D(A, B_k)$, and defined as

$$D(A, B_k) = \min\{D(\Gamma_{\mathrm{Ai}}, \Gamma_{B_k}) \mid i = 1, 2, 3, 4\}$$

Step 5: Array all the acquired distances $D(A, B_k)$ in increasing order. Output the first k corresponding images, i.e. the most similar k images to the query image A.

Comparing the image retrieval experiment results, we find that the V-moments constructed by functions in the first three groups (the V-moments of order 3 with $m, n = 0, 1, 2, \cdots, 7$) have the highest retrieval accuracy. In this case the feature vector is a 64-dimensional vector $\{\mu_{00}, \mu_{10}, \mu_{01}, \cdots, \mu_{77}\}$.

6. SHAPE RETRIEVAL EXPERIMENTS

In this section, we conduct a comprehensive performance comparison of the proposed algorithm with 5 classical moment algorithms, including 15th order Zernike moments (ZM), Hu invariant moments (HM), the 7th order orthogonal Fourier-Mellin moments (OFMM), the 11th Legender moments (LM), and 7th order geometric centroid moments (CM). In the experiments, the 3th order V-moments (VM) is used. The chosen orders in all methods are the best results order by large amounts of experiments.

Different algorithm may have different preprocessing method, the preprocessing of the algorithms used in this paper is as follows: method for Zernike moments and orthogonal Fourier-Mellin moments is the same: move the centroid of the image to the origin, draw a circumscribed circle around the target, further draw the outer tangential square, and cut out the image in the domain of the square, finally normalize it to 128×128 pixels; method for the V-moments, the geometric center moments and Legender moments is the same as the one described in section IV; for Hu moments the only thing need to do is moving the centroid of the image to the origin.

We adopt three indexes, Nearest Neighbor (NN), First-tier (FT) and Second-tier (ST), to evaluate the retrieval efficiency. In these three indexes, the retrieval accuracy is calculated by considering the first (not including the query image itself), the first m and the first $2m$ closest matches (including the query image itself), respectively, where m is the number of image in each class. For analysis of time efficiency, we list the average retrieval time per query, where the retrieval time is the time taken from input the query image to present the most similar $2m$ ones.

Our experiments are conducted on benchmark database MPEG-7-Shape- CE2 (CE2 for short), CE2 comprises images with complex shapes and textures, and multi closed boundaries, which is employed to test the recognition ability of region-based algorithm to complex images. CE2 is designed to test a region shape descriptor's behavior under different shape variations, and it has five sets (Sets A1, A2, A3, A4 and B). Set A1-A4 is used for testing scale invariance, rotation invariance, rotation and scale hybrid invariance, robustness to perspective transform, respectively. Set B is used for subjective similarity test.

The average retrieval accuracy and time efficiency are computed in our experiments. The experimental results are shown in Table I. The experimental results on sets A1, A3, and B show that the algorithm in this paper outperforms the classical methods compared. It means that this algorithm is preferable to recognize similarity transformed and visually similar images, the similarity transforms include scaling transform, rotation transform and TRS hybrid transform. It is well know that Zernike moment is currently one of the most effective image feature description methods, which has rotational invariance, but it is not robust to scaling distortion, so its result of experiment conducted on dataset Set A1 is not satisfactory. The V-moment algorithm proposed in this paper is robust to changes in scale although its robustness against rotation is slightly inferior to Zernike moment, so the experimental result on Set A1 for the V-moment is much better than those for other algorithms, and result on the dataset Set A3, which is used to evaluate robustness against scale and rotation, also shows the advantage of the algorithm in this paper. However, the experimental results on database A4 are not satisfying for the V-moments. It means that the new algorithm is not suitable for identify the similarity after perspective transform. It is not so important because perspective similarity is not our focus compared with scaling similarity and rotation similarity. The experimental results also show that the orthogonal Fourier-Mellin moments offer better performance to the perspective transformation in terms of robustness. In the aspect of time efficiency, the algorithm proposed in this paper is not as good as the Hu moments and Zernike moments, but outperforms the others. The main reason is the time of image preprocessing is too long by V-moments.

Table 1. Performance and efficiency comparison of different methods on the database mpen-7-shape-CE2 (%)

		VM **(the proposed)**	ZM	HM	OFMM	CM	LM
A 1	NN	**94**	90	41	86	62	87
	FT	**88**	71	34	70.2	47.8	75
	ST	**91.2**	77.4	36.4	72.8	56.8	82.2
	Time(s)	0.2245	0.0794	**0.0318**	0.9861	0.3933	3.8147
A 2	NN	99.29	**100**	78.57	93.57	90	99.29
	FT	95.61	**97.55**	68.98	80	75	93.98
	ST	96.12	**97.96**	82.35	80.41	83.67	95.41
	Time(s)	0.6024	0.1864	**0.0789**	2.9038	0.8949	11.5443
A 3	NN	**98.79**	97. 88	52.12	96.06	88.18	98.79
	FT	**95.87**	92.34	33.97	82.37	68.21	93.61
	ST	**97.13**	94.27	42.07	84.71	77.25	96.2
	Time(s)	1.4978	0.1807	**0.1261**	1.2132	1.7684	5.8045
A 4	NN	86.97	90.91	33.03	**93.94**	26.67	75.45
	FT	43.31	55.18	22.84	**63**	18.57	36.97
	ST	47.58	59.83	28.65	**68.95**	21.87	42.4
	Time(s)	1.5013	0.1745	**0.1085**	1.1643	1.8064	5.8543
B	NN	**73.46**	62.67	25.07	56.45	37.83	62.32
	FT	**29.15**	23.07	10.55	21.32	20.26	24.44
	ST	**42.12**	33.57	16.96	29.36	32.15	35.6
	Time(s)	0.2285	0.0929	**0.0345**	1.1696	0.4322	3.9612

7. CONCLUSION

In this paper, a new class of moment functions called V-moments is defined using an orthogonal function system, V-system. By applying them on feature extraction of images, a new algorithm for image retrieval thereby is proposed. Since the V-system contains a lot of discontinuous basis functions, it has more advantage in extracting features of multi-boundary image. In general, an image can be approximately characterized by its boundary, however, when the boundary of an image is very complex, its boundary extraction becomes a much harder problem, so the boundary-based feature is generally not reliable. Complex image retrieval is a difficult problem in image retrieval. As the V-moment takes into account all image regions, an image whether it is simple or complex can be nearly reconstructed by it, therefore the V-moment is more reliable for complex image retrieval. In this paper, a plurality of evaluation indices is adopted, and performance of the V-moments is compared with that of several classical moment methods. In these experiments, the V-moments have acquired high retrieval accuracy rate and satisfactory time efficiency, which demonstrates that the algorithm proposed is practical and efficient.

ACKNOWLEDGEMENT

This research is supported by National Key Basic Research Project of China under Grant No. 2011CB302400, National Natural Science Foundation of China under Grant No. 61272026, Beijing Natural Science Foundation Program and Scientific Research Key Program of Beijing Municipal Commission of Education under Grant No.KZ201210009011.

REFERENCES

[1] Ming-Kuei Hu, 1962. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, Vol.8, No. 2, pp 179–187.

[2] Shan Li, et al, 2009. Complex Zernike Moments Features for Shape-Based Image Retrieval, *IEEE Transactions on Syatems, Man, and Cybernetics—Part A: Syatems and Humans*, Vol.39, No. 1, pp 227-237.

[3] Li-Qiang Guo, Ming Zhu, 2011. Quaternion Fourier–Mellin moments for color images. Pattern recognition, Vol.44, No. 2, pp 187-195.

[4] Hui Zhang, et al, 2010. Blurred image recognition by Legendre moment invariants, *IEEE Transactions on Image Processing*, Vol.19, No. 3, pp. 596 – 611.

[5] Innchyn Her, et al, 2011. A Hybrid Trademark Retrieval System Using Four-Gray-Level Zernike Moments and Image Compactness Indices. *International Journal of Image Processing*, Vol.4, No. 6, pp 631-646.

[6] Naif Alajlan, et al, 2006. Multi-object image retrieval based on shape and topology. *Signal Processing: Image Communication*, Vol.21, No.10, pp 904–918.

[7] Heng Qi, et al, 2010. An effective solution for trademark image retrieval by combining shape description and feature matching. *Pattern Recognition*, Vol.43, No.6, pp 2017–2027.

[8] Chandan Singh, Pooja, 2012. An effective image retrieval using the fusion of global and local transforms based features. *Optics & Laser Technology*, Vol.44, No.7, pp 2249–2259.

[9] Anjali Goyal, Ekta Walia, 2012. Variants of dense descriptors and Zernike moments as features for accurate shape-based image retrieval. *Signal, Image and Video Processing*, Vol.6, No.1, pp 1-17.

[10] Akriti Nigam, et al, 2011. Content based Trademark Retrieval by Integrating Shape with Color and Texture Information. *International Journal of Computer Applications*, Vol.22, No.7, pp 40-45.

[11] Chia-Hung Wei, et al, 2009. Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, Vol.42, No.3, pp 386-394.

[12] NacéraLaiche, et al, 2014. Curve normalization for shape retrieval. *Signal Processing: Image Communication*, Vol. 29, No.4, pp556–571

[13] Qi Jia, et al, 2014. A new geometric descriptor for symbols with affine deformations. *Pattern Recognition Letters*, Vol.40, pp128-135

[14] Ruixia Song, et al, 2007. A New Class of Complete Orthogonal V-system and Its Applications. *Communication on Pure and Applied Analysis*, Vol.6, No.3, pp 853 –871.

[15] Chao Huang, et al, 2012. A new class of multi-wavelet bases: V-system. *Acta Mathematica Sinica, English Series*, Vol.28, No.1, pp 105-120.

THICK MESH CONSTRUCTION FROM SCATTERED CLOUD

Thierry Bay[1] and Romain Raffin[2]
[1]*LAMAV (FR CNRS 2956) - EA 4015, University of Valenciennes - ISTV2 - Mont Houy, 50313 Valenciennes Cedex 9*
[2]*Aix-Marseille University, LSIS UMR 7296, Polytech - Case 925, 163, Avenue de Luminy, 13288 Marseille Cedex 9*

ABSTRACT

In a medical context, we reconstruct a surface from a sparse point cloud. The output of the 3D processing must be the input in a mechanical solver (FEM) to characterize the stress applied to the organs through the constitutive laws. To be as generic as possible, we construct a parametric surface. This surface is defined by a single B-spline patch, providing an easy way to parameterize, as well as a wish of the FEM team which operates on the resulting hexahedral model after the addition of the thickness. We present a method to automatically construct discrete offsets layers based on the physiological thickness of each organ.

KEYWORDS

3D reconstruction, parametric surface, offset.

1. INTRODUCTION

This paper addresses the problem of a digital reconstruction of the pelvic floor and the organs in the caudal area of the basin: the bladder, the rectum, the uterus and the vagina. An “anatomically medically realistic” geometric modelling of all these organs is required, to simulate *a posteriori* their behaviour through the constitutive laws. Many simulators have been introduced on the market since the second half of last century (cf. [Zhong et al., 2012], [Zhu and Gu, 2012], [Patete et al., 2012], [Niroomandi et al., 2008], [Sela et al., 2007]). Two different orientations emerge from the current works: the first ones are based on the principles of continuum mechanics (cf. [Wang et al., 2007], [Müller et al., 2004]), and the second ones are not (cf. [Müller et al., 2005], [Picinbono et al., 2002]). However, the geometric modelling of the organs is generally underestimated in order to meet real time constraints at the expense of in vivo behaviours, or to make up for non-realistic movements only with physical parameters.

Therefore, we propose a solution coming up with the underlying geometric model by defining criteria needed for second generation simulators (cf. [Satava, 1996], [Sierra et al., 2006], [Schwartz et al., 2005]). A geometric support is built for each organ by a two-step process: the construction of the surface mesh and the addition of a thickness. We prefer to remain as faithful as possible (with respect to the data) to physiological reality and we relax the real-time constraint (around one minute of processing). The surface mesh is defined with a 0-genus parametric surface, making possible to work with a global unique parameterization. Usually, literature about the parametric representation is more focused on triangular patches to work on arbitrary topologies (cf. [Nagata, 2005], [Walton and Meek, 1996], [Hahmann and Bonneau, 2003], [Funfzig et al., 2008]). However, we can limit our work to 0-genus topology since the input objects do not present holes (apart from extremities if junctions are needed). Besides, we bypass the vertex consistency problem arising when we need to deal with continuity between several triangular patches with Bézier triangles [Boschiroli, 2012]: a linear system has to be solved to each vertex, whose the size depends on the 1-neighborhood; with a dense mesh, this work is time consuming.

The shape parameters used in the parametric representation differ according to the nature of the surface. Bézier, Bspline, NURBS or T-spline are possible formulations for a 3D model (cf. [Piegl and Tiller, 1997, Wang et al., 2011]). They are well adapted for open shapes, but specific care - described below - is necessary to impose a continuity constraint everywhere for closed 3D models.

About the offset approach, the State-Of-the-Art written by [Maekawa, 1999] puts forward the major interrogations related to the offset methods. Regarding our method, we focus on the following questions: "*how to approach the offset-surface*", and "*how to get rid off the self-intersections*"? To answer these questions, two branches of offset methods are put forward in the literature: volumetric and surface based approaches. Volumetric methods embed the input geometry into a grid of voxels. That permits to localize and to interact with a generally implicit formulation of the offset but the final resolution is strongly dependent of the initial grid resolution. We can cite [Varadhan and Manocha, 2004, Chen et al., 2005, Breen et al., 1998, Liu and Wang, 2011] where Marching-Cube based methods, or matching between the initial point cloud and its offset are performed. *A contrario*, in surface based methods, two types of method, which no longer consider volumetric grids are common: the first one is based on the implicit representation with radial basis functions, and the second one uses the parametric representation. Among the implicit methods, [Liu et al., 2009] propose an algorithm to simultaneously reconstruct the surface approaching an oriented point cloud and its offset. To handle the case of sparse data with holes and a large offset distance, a RBF with compact supports is used (cf. [Ohtake et al., 2005]). This approach is effective, but does not satisfy constraints of parameterization for an accurate location on the surface. Moreover, in detailed areas of the original model, artefacts can be created when the offset-distance is too large (which sometimes may occur with organs like the uterus). Among the parametric offsets, [Kulczycka and Nachman, 2002] divide the approaches into two families belonging to C^1 ([Farouki, 1986], [Hoschek et al., 1989]), or C^2 continuous methods [Tiller and Hanson, 1984]. Some methods consist in directly offsetting the sampling of the surface and then fit the offset cloud, while the others offset directly the control network. Nevertheless, the management of self-intersections is not included in their basic processes, which requires a specific care. Either the control points are repositioned to reduce the local curvature, or parametric restrictions are applied to prevent the self-intersections. A common drawback of these methods is their failure to detect tiny self-intersections (cf. [Sun et al, 2004], [Seong et al, 2006]).

According to our project's goal, the method must provide a hexahedral mesh. Moreover, a limitation of the current offset models has not been mentioned, and directly concerns the realism of modelling. The works in the literature relied on the creation of a thick membrane with a uniform thickness. However, for shapes with at least one opening such as the uterus and the rectum, geometric models must include the ability to obtain a smaller thickness at the ends, which can therefore not be the same values in the rest of the 3D model. Current approaches do not take into account this factor, which is still challenging to obtain. To lead to more realistic shapes from a medical point of view, a discrete method is defined by use of a median axis. An estimate of the cavity is computed, instead of computing the offset from the outer shell. By defining a quadrangular tessellation for the outer surface, hexahedral meshes are finally obtained.

2. SURFACE FITTING

In a segmentation process, physician experts extract a scattered set of spatial points belonging to the surface of each organ of interest. The organs considered herein are 0-genus surfaces of approximately 15x15 cm, with a thickness of 3-7 cm. From each point cloud, the fitting process constructs the approximating parametric surface describing the outer shell of an organ. In order to increase the method reliability, a two-step initialization process is used. Firstly, a PCA creates indicators on the global shape to get the two furthermost points of the elongated organ. From these two points, the associated bounding box is used to determine the principal axis of an ellipsoid surrounding the point cloud. We perform then an approximation based on the minimization of distances between the 3D point cloud and a Bézier curve fixed to the previous two points to obtain a curvilinear median axis. Secondly, this axis is used to construct the control network of the initial approximating surface by extruding a set of circular closed polylines of control points (whose radius is given by the distance between the axis and the point cloud) orthogonally to the curve (cf. [Bay et al., 2012a]). The curvilinear axis is therefore a guideline to position a generalized cylinder along its direction that roughly fits the set of points. This rough surface is a semi-periodic B-spline patch: periodic in one parametric direction and closed in the second direction to meet continuity constraints over the whole surface.

From the parametric surface, the energy function has to quantify the fitting errors. In most of B-spline based methods, an alternate scheme is applied: parameterization of the data by orthogonal projections on the surface, and steepest descent to move the control points. However, the idea behind the two-termed energy is to stabilize the parameterization step, which can lead to several problems (false/double-projections) if the initial surface is far from the point cloud to be fitted. In such cases, parts of the point cloud or of the surface are not considered during the fitting, and ripples may therefore occur. Even though a regularization step could solve these problems, stability is increased without these extra calculations unrelated to the fitting term.

For the sake of clarification, the principle is described for a 2D parametric curve fitted on a scattered point cloud. The following two-termed energy is used (see Figure 1). The first energy associates the closest sampled point in terms of Euclidean distance to each data point, and the second energy associates the closest data point in terms of Euclidean distance to each sampled point; the error function measures a co-distance based energy between the data points and the curve C (cf. Equation 1). The next step consists in reducing the energy function; it is obtained with the alternate iterative Hoschek-like method by moving the control points.

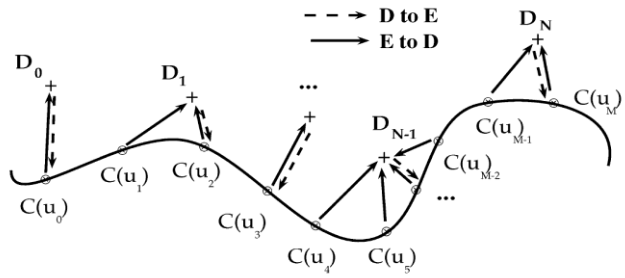


Figure 1. Principles of two distance functions for the curve C

More details about the energy function are provided below. The error function is separated into two sub-functions (cf. Equation 1): the whole dataset and the entire curve are considered to fit the main shape of the 3D model. In this way, the system is more efficient to reduce ripples (important for surface fittings). The first one, noted $f_{D\to E}$ in Equation 1, reflects the dissimilarity of D with respect to the sampling E of C (their connections will be collected in U) and the second one $f_{E\to D}$. The next equation reflects the one of E over D (in the same way, their connections will be collected in a set G):

$$f(P,U,G) = f_{D\to E}(P,U) + f_{E\to D}(P,G) \qquad (1)$$

The sub-function $f_{D \to E}$ maps each point of D with the closest sampled point evaluated on the curve in terms of Euclidean distance:

$$f_{D\to E}(P,U) = \frac{1}{N+1}\sum_{s=0}^{N}\left\|D_s - C(u_{D_s})\right\|^2 \qquad (2)$$

with: $\left\|D_s - C(u_{D_s})\right\|^2 = \min_j \left\|D_s - C(u_{D_j})\right\|^2$. The closest evaluated point of D_s is noted $C(u_{D_s})$, with u_{D_s} the parameter of D_s on the curve C (by abuse of notations, and not to be confused with $C(u_s)$, the s^{th} evaluation of the curve in the parametric domain). On the same principle, the second sub-function $f_{E\to D}$ maps each evaluated point on the curve to the closest data point of the dataset D in terms of Euclidean distance:

$$f_{E\to D}(P,C) = \frac{1}{M+1}\sum_{k=0}^{M}\left\|D_{u_k} - C(u_k)\right\|^2 \qquad (3)$$

with: $\left\|D_{u_k} - C(u_k)\right\|^2 = \min_l \left\|D_{u_l} - C(u_k)\right\|^2$.

We assume that the vertices of the input dataset are regularly distributed, so we discretize the fitting surface with the same number of vertices. The use of two distance functions does not imply projection unlike usual methods. Furthermore, averaging the distance from on set to the over raises the robustness of the algorithm in noisy points set. The same approach based on two distances is used for the surface to move the control points and approximate the data set.

Control over the surface is made by some improvements added to the global process we described above. In order to improve the smoothness if noisy datasets are used, the surface curvature is minimized by applying a Hadenfeld-like method ([Hadenfel, 1995]). A second improvement refers to the details preservation, made possible by the knot insertion where the errors to the data are the most important. However, the topological rule for quadrangular network limits the process. Some results of our method are shown hereafter.

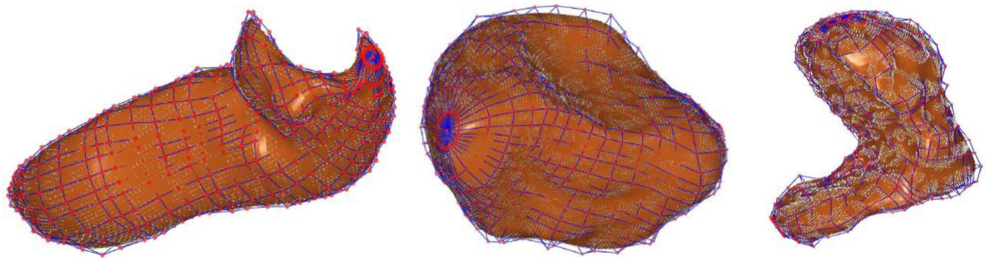


Figure 2. Fitting results for each organ

Figure 2 shows the fitting of a set of organs (left rectum, middle bladder, and right uterus/vagina). Errors and behaviours of an energy function and the global fitting function are detailed in Table 1.

Table 1. Numerical results of the fitting process

	$E_{\max}$ (mm)	EQM (mm) $D \to E$	Time (s)
Rectum	0.942	0.158	37
Bladder	0.988	0.153	49
Uterus/vagina	3.26	1.02	108

The surface approximation stops when the error (average distances between data set and surface) is below the error a human operator can do in the image segmentation process (around 1mm). With the uterus, the value is greater than the 1mm threshold due to outliers inside the point cloud.

3. OFFSET CONSTRUCTION

In section 2, we have described the process developed to get the outmost shell of the organs. From this layer, we have to obtain the inner wall of the cavity – referenced later as the "offset" – to construct the hull of our "3D thick mesh". The mesh of the inner wall is constructed from a discretization of the approximating closed surface, extended towards the interior of the surface with an "offset method". It uses physiological values to determine the thickness distance. As we manipulate a continuous parametric closed surface for the external shell, we can use the global parameterization to cover the surface. The discretization is managed so that quadrilateral meshes are obtained, leading to hexahedral volume meshes with its offset. Contrary to other approaches [Hoschek et al., 1989, Tiller and Hanson, 1984, Kulczycka and Nachman, 2002] where the parametric surface guides the offset construction, our choice is motivated by the fact that the data describing the outer mesh contains errors we do not want to transfer inside. Our construction is based on medical data that estimate a longilineal cavity, and not on a cavity computed from the external shell. For organs such as the uterus, the cavity is so narrow that the uterine walls are in contact. The construction of this cavity from the outer shell would lead to stability problems, since more and more irregularities - or self-intersections - would be produced as the thickness increases. The following approach bypasses this problem of stability by a two-step process (see Figure 3). In practical terms, we:

1. Position an initial cylindrical mesh of small radius along the median axis of the outer mesh.
2. Inflate the inner mesh outward guided by the connections between the discretization of the external parametric surface and the cylindrical internal mesh.

Let C be the Bézier curve, corresponding to the median axis, and D be the dataset. By construction, the extremities of the curve belong to D, creating an axis that describes the main direction. Let $\{M^0_{i,j}\}^{M_1,M_2}_{i,j=0}$ be the external mesh (sampling of S), and $\{M^1_{i,j}\}_{n,m}$ be the internal mesh. The mesh M^1 is initialized by a succession of m+1 circles. Each circle is built with n+1 points, is orthogonal to the median axis, has a 1 mm radius (minimal threshold chosen for the cavity), and their centre belongs to the median axis. To position these circles, their centres are computed as follows. Let $\{c_j\}_{j=0}$ be the centres, for 0<j<m we have:

$$c_j = C\left(\arg\min_s \left(\left\| \frac{\sum_{i=0}^{n+1} M^1_{i,j}}{n+1} - C(s) \right\|^2\right)\right)$$

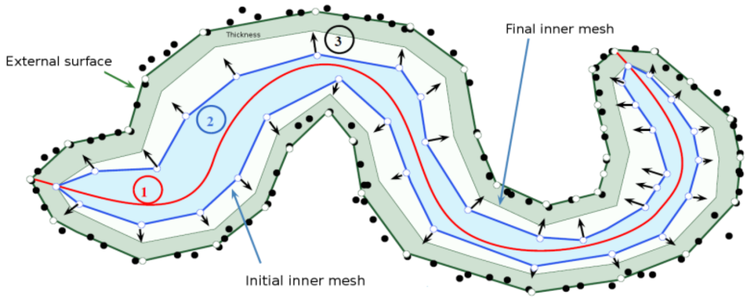


Figure 3. Offset construction, inflating an inner mesh

A parameter s_j is then associated to each centre c_j, such as $C(s_j) = c_j$. A special case concerns the extremities c_0 and c_m, otherwise the meshes M^0 and M^1 would be in contact. Indeed, the sets $\{M^1_{i,0}\}^{i=0}_n$ and $\{M^1_{i,m}\}^{i=0}_n$ are respectively merged with extremities at the beginning. An offset is then applied to move them away these points at a distance d_*. This value is determined such that $d_* \le d_0$, with d_0 the offset-distance. This modification is reflected in the nearby centres, since they need to be repositioned uniformly. The orientation of the sets $\{M^1_{.,j}\}$ for each circle has to be managed carefully. Otherwise, the connections between M^0 and M^1 would intersect. For each circle, the points are uniformly distributed. The discretization of two neighbour circles is modified so that the shortest path between the two circles is guaranteed. The set of remaining points $\{M^1_{i,j}\}$ for the j^{th} circle is then oriented by minimizing the sum of distances between $\{M^1_{i,j}\}$ and $\{M^1_{i+1,j}\}$. Finally, the evolution of the system is governed by the satisfaction of a set of constraints to authorize or not the displacement of a point. The displacement of $M^1_{i,j}$ is rejected if quadrangles associated with $M^1_{i,j}$ are not crossed, if the segment $\left[M^0_{i,j} M^1_{i,j}\right]$ does not cross quadrangle of M^1, or if the distance from $M^1_{i,j}$ to its projection on S is lower than d_0. The process ends as soon as each particle is stationary. The new position at time $t+\Delta t$ for $M^1_{i,j}$ is computed as follows, with $\Delta x > 0$ the displacement step:

$$M^1_{i,j}(t+\Delta t) = M^1_{i,j}(t) + \Delta x \frac{\overrightarrow{M^1_{i,j}(t) M^0_{i,j}(t)}}{\left\| M^1_{i,j}(t) M^0_{i,j}(t) \right\|}$$

Table 2 presents some results of our method of offsets construction compared to a mass-spring method [Bay et al, 2012b]. It shows that the method is rapid and does not raise self-intersection in the hexahedral mesh. We choose this set of points because its offset-distance is more complex to handle. The thickness is specified by a unique value for the two organs even though one is slender while the other is rounded.

ISBN: 978-989-8533-38-8 © 2015

Table 2. Numerical results of an offset built for the couple uterus/vagina (in mm)

	Min. Error	Max. error	Average error	Self-intersection?	Time (s)
Mass-spring method	16.5	26.8	20.7	Yes	59
Our method	3.4	19.4	18.8	No	12

As we can see on Figure 4, on the left with the mass-spring method, the offset for the uterus/vagina object is self intersecting (dark interior parts). The middle image given by our method is free of self-intersection, despite a stretched area in the vagina part (in the bottom).

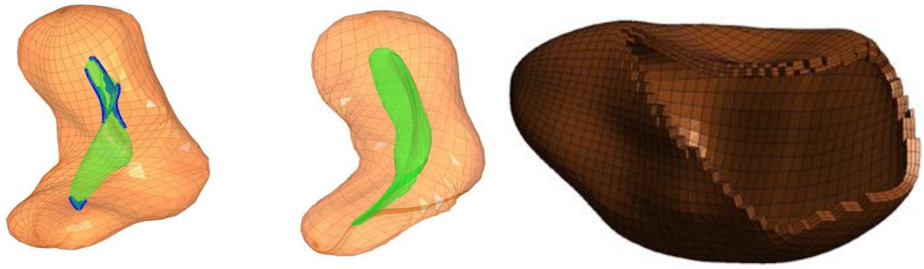


Figure 4. Uterus/vagina results of offset meshes (left-middle), offset layers of a bladder (right)

4. OFFSET LAYERS

In order to provide flexibility to FEM experts, several layers are constructed. Topology conservation between the outer and inner meshes provides appropriate connections to get hexahedral meshes.
Only two physical constraints were given regarding the meshes:

- To provide an hexahedral mesh per dataset.
- To be as accurate as possible regarding the medical data, i.e. use the medical knowledge (the thickness) we have at our disposal.

However, in the interests to obtain quality elements required by the FEM software, adaptations regarding the shape deviation of each hexahedron are necessary. For the moment, the processing avoids dealing with a too much prominent dimension for each hexahedron by applying a linear interpolation between the two inner and outer meshes. The volume mesh for the bladder puts forward the individual layers of this organ (Fig. 4-right).

5. CONCLUSION AND FUTURE WORKS

This work wishes to provide a new approach to build a thick surface from a scattered point cloud describing an elongated form. We propose a method to control directly the outer layer with the parametric representation, and to get closer to physiological reality with the thickness by moving away from traditional methods and approaches. The method for the reconstruction of the surface from 3D models has indeed proved its efficiency with organs. With the connection graphs, several problems have been avoided or bypassed, such as the false orthogonal projections, or continuity management which would have been necessary with curved triangular patches. Furthermore, the co-distance based energy has yielded a more stable method than referenced publications, by avoiding the systematic use of an empiric regularizing term. Finally, the parametric surface makes the application of remeshing operators possible without losing data. Additionally, we control the tessellation of the surface, according to physiological, geometrical, mechanical constraints. The model is also able to create holes by parameters restrictions. These features cannot be handled by points or mesh reconstruction algorithms. For the construction of the offset, the literature highlights two approaches often used: the implicit representation and volumetric techniques. We wanted to develop a method allowing both to keep direct control over the reconstruction, and to create an offset mesh whose the resolution would not depend on a grid of voxels. We have adopted a reverse orientation compared to current approaches, by estimating the cavity rather than computing the inner mesh from the outer layer. The method based on the median axis has helped to manage the self-intersections (stopping the process in case of detection), allowing the use of variable thicknesses. In addition, the transfer of errors coming from the surface fitting was avoided. However, since the medical specific knowledge is lacking, the accuracy of the physiological cavity is unknown and can not be quantitatively validated. Nevertheless, a hexahedral conform non-mixed mesh without overlap can be provided to the physical modelling for finite element simulations. One limitation of this offset method remains its use case, since the input shape has to have a principal direction.

In order to apply an offset operator in a more general context, a new purely geometric approach could be investigated for triangular meshes. After the calculation of each discrete normal, the offset of each point would be obtained. The second step would be to construct a generalized offset. In other words, a local modification of the offset distance or of the angular deviation from the normal vector could be done to manage the self-intersections. An energy function will modify the normal vector (its norm or its deviation) to correct this problem.

REFERENCES

Bay, T. et al, 2012a. Geometric modeling of pelvic organs with thickness. *In 3DIP*, Burlingame, USA, Vol. 8290, pp. 82900I 1-82900I 14.

Bay, T. et al, 2012b. Geometric modeling of pelvic organs with a discrete offset approach. *In 20th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. Plzen, pp. 27-30.

Boschiroli M. et al, 2012. G1 rational blend interpolatory schemes: A comparative study. *In Graphical Models*, Vol. 74, pp. 29-49.

Breen D. E. et al, 1998. 3D Scan Conversion of CSG Models into Distance Volumes. *In Symposium on Volume Visualization*, Durham, USA, pp. 7-14.

Cazals, F. and Giesen, J., 2006. Delaunay triangulation based surface reconstruction: ideas and algorithms. *In Effective Computational Geometry for Curves and Surfaces*, pp. 231-273.

Chen Y. et al, 2005. Filleting and rounding using a point-based method. *In Proceedings of ASME international design engineering technical conferences*, Long Beach, USA, Vol. 2, pp. 533-542.

Delingette, H. and Ayache, N., 2004. Soft tissue modeling for surgery simulation. *In Computational Models for the Human Body, N. Ayache*, Vol. 12, pp. 453-550.

Farouki, R., 1986. The approximation of non-degenerate offset surfaces. *In CAGD*, Vol. 3, pp. 15-43.

Hadenfeld, J., 1995. Local energy fairing of b-spline surfaces. *In Mathematical Methods for Curves and Surfaces*, pp. 203-212.

Hahmann, S. and Bonneau, G. P., 2003. Polynomial surfaces interpolating arbitrary triangulations. *In IEEE Transactions on Visualization and Computer Graphics*, Vol. 9, No 1, pp. 99-109.

He Y. et al, 2006. Automatic shape control of triangular b-splines of arbitrary topology. *In Journal of Computer Science and Technology*, Vol. 21, No 2, pp. 232-237.

Hoppe H. et al, 1992 . Surface reconstruction from unorganized points. *In Computer graphics*, Vol. 26:2, pp. 71-78.

Hoschek, J., 1988. Intrinsic parametrization for approximation. *In Computer Aided Geometric Design*, Vol. 5, pp. 27-31.

Hoschek, J. et al, 1989. Optimal approximate conversion of spline surfaces. *In Computer Aided Geometric Design*, Vol. 6, pp. 293-306.

Ju T. et al., 2002. Dual contouring of Hermite data. *In ACM Transactions on Graphics*, Vol. 21, pp. 339-§346.

Kulczycka, M. A. and Nachman, L. J., 2002. Qualitative and quantitative comparisons of B-spline offset surface approximation methods. *In Computer-Aided Design*, Vol. 34, No 1, pp. 19-26.

Kumar G. V. V. et al, 2002. Computing non-self-intersecting offsets of NURBS surfaces. *In Computer-Aided Design*, Vol. 34, pp. 209-228.

Liu, S. and Wang, C. C. L., 2009. Duplex fitting of zero-level and offset surfaces. *In Computer-Aided Design*, Vol. 41, No 4, pp. 268-281.

Liu, S. and Wang, C. C. L., 2011. Fast Intersection-Free Offset Surface Generation From Free-form Models With Triangular Meshes. *In IEEE Transactions on Automation Science and Engineering*, Vol. 8, No 2, pp. 347-360.

Lorensen, W. E. and Cline, H. E., 1987. Marching cubes: A high resolution 3d surface construction algorithm. *In Computer Graphics*, Vol. 21, No 4, pp. 163-169.

Müller et al., 2005. Meshless deformations based on shape matching. *In ACM Transactions on Graphics*, Vol. 24, No 3, pp. 471-478.

Nagata T., 2005. Simple local interpolation of surfaces using normal vector. *In Computer Aided Geometric Design*, Vol. 22, No 4, pp. 327-347.

Niroomandi et al., 2008. Real-time deformable models of non-linear tissues by model reduction techniques. *In Computer Methods and Programs in Biomedicine*, Vol. 91, No 3, pp. 223-231.

Ohtake Y., 2005. 3D scattered data interpolation and approximation with multilevel compactly supported RBFs. *In Graphical Models*, Vol. 67, No 3, pp. 150-165.

Patete et al., 2012. A multi-tissue mass-spring model for computer assisted breast surgery. *In Medical Engineering & Physics*, Vol. 35, No 1, pp. 47-53.

Pekerman D. et al, 2008. Self-intersection detection and elimination in free-form curves and surfaces. *In Computer-Aided Design*, Vol. 40, pp. 150§-159.

Picinbono et al., 2002. Improving realism of a surgery simulator: linear anisotropic elasticity, complex interactions and force extrapolation. *In Journal of Visualization and Computer Animation*, Vol. 13, No 3, pp. 147-167.

Piegl L. et al, 1997. *The NURBS book (2nd edition)*. Springer-Verlag, New York, USA.

Randrianarivony, M. and Brunnet, G., 2002. Approximation by NURBS curves with free knots. *In Vision, Modeling and Visualization.* Erlangen, Germany, pp. 195-201.

Satava R., 1996. Medical virtual reality: the current status of the future. *In Medicine Meets Virtual Reality: Health Care in the Information Age*, Vol. 29, pp. 100-106.

Saux, E. and Daniel, M., 2003. An improved Hoschek intrinsic parametrization. *In Computer Aided Geometric Design*, Vol. 20, No 8, pp. 513-521.

Schwartz et al., 2005. Modelling liver tissue properties using a non-linear viscoelastic model for surgery simulation. *In Medical Image Analysis*, Vol. 9, No 2, pp. 103-112.

Seong J. K., 2006. Trimming Local and Global Self-intersections in Offset Curves/Surfaces Using Distance Maps. *In Computer Aided Design*, Vol. 38, No 3, pp. 183-193.

Sharf A. et al, 2006. Competing fronts for coarse-to-fine surface reconstruction. *In Eurographics Computer Graphics Forum*. Vienna, Austria, pp. 389-398.

Sun Y. et al, 2004. Modifying free-formed NURBS curves and surfaces for offsetting without local self-intersection. *In Computer-Aided Design*, Vol. 36, pp. 1161-1169.

Tiller, W. and Hanson, E., 1984. Offsets of Two-Dimensional Profiles. *In IEEE - Computer Graphics and Applications*, Vol. 4, pp. 36-46.

Varadhan, G. and Manocha, D., 2004. Accurate Minkowski Sum Approximation of Polyhedral Models. *In Graphical Models*, Vol. 68, No 4, pp. 343-355.

Walton, D.J. and Meek, D.S., 1996. A triangular G1 patch from boundary curve. *In Computer Aided Design*, Vol. 28, No 2, pp. 113-123.

Wang P., 2007. Virtual reality simulation of surgery with haptic feedback based on the boundary element method. *In Computers & Structures*, Vol 85, No 7-8, pp. 331-339.

Wang W. et al, 2011. Converting an unstructured quadrilateral mesh to a standard T-spline surface. *In Computational Mechanics*, pp. 477-498.

Zhang Y. et al, 2012. Solid T-spline construction from boundary representations for genus-zero geometry. *In Computer Methods in Applied Mechanics and Engineering*, Vol. 249-252, pp. 185-197.

Zhong et al., 2012. Soft tissue deformation with reaction-diffusion process for surgery simulation. *In Journal of Visual Languages & Computing*, Vol. 23, No 1, pp. 1-12.

Zhu and Gu, 2012. A hybrid deformable model for real-time surgical simulation. *In Computerized Medical Imaging and Graphics*, Vol. 36, No 5, pp. 356-365.

Hadenfeld J., 1995. Local energy fairing of B-spline surfaces. *In Mathematical Methods for Curves and Surfaces*, pp. 203-212.

Fünfzig C. et al, 2008. PNG1 triangles for tangent plane continuous surfaces on the GPU. *In Proceedings of Graphics Interface*, pp. 219-226, Toronto, Canada.

Jaillet F. et al., 1997. Periodic B-spline surface skinning of anatomic shapes. *In 9th Canadian Conference on Computational Geometry*. Kingston, Canada, pp. 119-210.

Kazhdan M. et al, 2006. Poisson surface reconstruction. *Proceedings of the fourth Eurographics symposium on Geometry processing*. Aire-la-Ville, Switzerland, pp. 61–70.

Müller et al., 2004. Point based animation of elastic, plastic and melting objects. *In Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Aire-la-Ville, Switzerland, pp. 141-151.

APPLIED SEQUENTIAL-SEARCH ALGORITHM FOR COMPRESSION-ENCRYPTION OF HIGH-RESOLUTION STRUCTURED LIGHT 3D DATA

Mohammed M. Siddeq and Marcos A. Rodrigues
GMPR-Geometric Modelling and Pattern Recognition Research Group, Sheffield Hallam University, Sheffield, UK

ABSTRACT

Anew image compression algorithm is proposed and demonstrated in the context of structured light 3D reconstruction. Structured light images contain patterns of light, which are captured by the sensor at very high resolution. The algorithm steps involvea two level Discrete Wavelet Transformation (DWT) followed by a Discrete Cosine Transformation (DCT) to generate a DC-Column andan MA-Matrix (Multi-Array Matrix). The MA-Matrix is then partitioned into blocks and a minimization algorithm codes each block followed by arithmetic coding. At decompression stage anewproposedalgorithm, Sequential-Search Algorithm (SS-Algorithm) is used to estimate the MA-Matrix. Thereafter, all decompressed DC-Columns arecombined withthe MA-Matrix followed by inverse DCT and inverse DWT. The effectiveness of the algorithm is demonstrated within a 3D reconstruction scenario from structured light images.

KEYWORDS

Discrete Wavelet Transform; Discrete Cosine Transform; Minimize Matrix Size Algorithm; Sequential Search Algorithm;Structured Light; 3D reconstruction.

1. INTRODUCTION

In today's highly computerized and interconnected world, security of digital images/video has become increasingly more significant in applications such as pay preview TV, confidential video conferencing, medical imaging and industrial or military imaging systems. Many different image encryption techniques have been proposed in the literature. They include Bit Recirculation Image Encryption, Infinite Series Convergence method [3], Fuzzy PN code based Color Image Encryption method, Combinational Permutation method, Magnitude and Phase Manipulation method, SCAN based methods and Chaos based methods. Further, image encryption based on phase encoding by means of a fringe pattern uses cosine function, which adds to its argument the image to be phase encrypted [4,11]. Since these methods manipulate an entire data set without any presumption about compression at later time, the secure transmission of image has become more costly in terms of time, bandwidth and complexity [16]. Thus, users pay a price for security proportional to their desired level of security. Further, the use of compression after encryption fails to exploit the spatial and psycho-visual redundancies efficiently as the encryption of an uncompressed image removes intelligibility from the original image and hence incurring compression penalties. This results in a tradeoff between the competing requirements of encryption and compression [20].Here a further requirement is introduced concerning the compression of 3D data. Rodrigues [13] demonstrated that while geometry and connectivity of a 3D mesh can be tackled by a number of techniques such as high degree polynomial interpolation or partial differential equations [14] the issue of efficient compression of 2D images both for 3D reconstruction and texture mapping for structured light 3D applications has not yet been addressed. Moreover, in many applications, it is necessary to transmit 3D models over the Internet to share CAD/CAM models with e-commerce customers, to update content for entertainment applications, or to support collaborative design, analysis, and display of engineering, medical, and scientific datasets. Bandwidth imposes hard limits on the amount of data transmission and, together with storage costs, limit the complexity of the 3D models that can be transmitted over the Internet and other networked environments [12,15].

ISBN: 978-989-8533-38-8 © 2015

The focus of this paper is on compression of structured light images. Such images contain patterns of light allowing 3D reconstruction. A possible scenario would be a surface patch compressed as a 2D image together with 3D calibration parameters, transmitted over a network and remotely reconstructed (geometry, connectivity and texture map) at the receiving end with the same resolution as the original data. The widespread integration of 3D models in different fields motivates the need to be able to store, index, classify, and retrieve 3D objects automatically and efficiently.

Siddeq and Rodrigues [12] proposed 2D image compression methods based on high-frequency sub-bands compressed by the Minimize-Matrix-Size Algorithm (MMS) and decompressed by the Limited-Sequential Search Algorithm (LSS). The advantages are high compression ratios with high-resolution 3D reconstruction. However, the complexity of the algorithm means very large execution times that could be in the order of minutes. A new algorithm was proposed [13] using JPEG, with decompression by a parallel SS-Algorithm. The execution time was reduced to a few seconds with higher compression ratios. Recently, a novel algorithm was proposed for decompression of DCT coefficients called Fast Matching Search (FMS), which reduced execution time to milliseconds. Further, the FMS algorithm was applied to frequency sub-bands of DWT followed by DCT [23]. In this research we introduce a new method for compression and encryption by partition the DCT coefficients into blocks and applying the MMS-Algorithm on each block of pixels. Each block generates a unique key at compression stage. This key can be seen as a symmetric encryption key, as without the key the block cannot be decoded.

2. THE PROPOSED COMPRESSION-ENCRYPTION ALGORITHM

The proposed algorithm uses two transformations:a two level DWT converts an image into seven frequency sub-bands. The low-frequency sub-band is divided into 2x2 non-overlapped blocks and a DCT is applied to each block. The DCT is very important to increase high-frequency domains by converting LL2 into DC-coefficients and AC-coefficients (DC-Column and MA-Matrix).The Minimize-Matrix-Size Algorithm is applied to MA-Matrix for encryption and then subject to arithmetic coding together with DC-Column as depicted in Figure1.

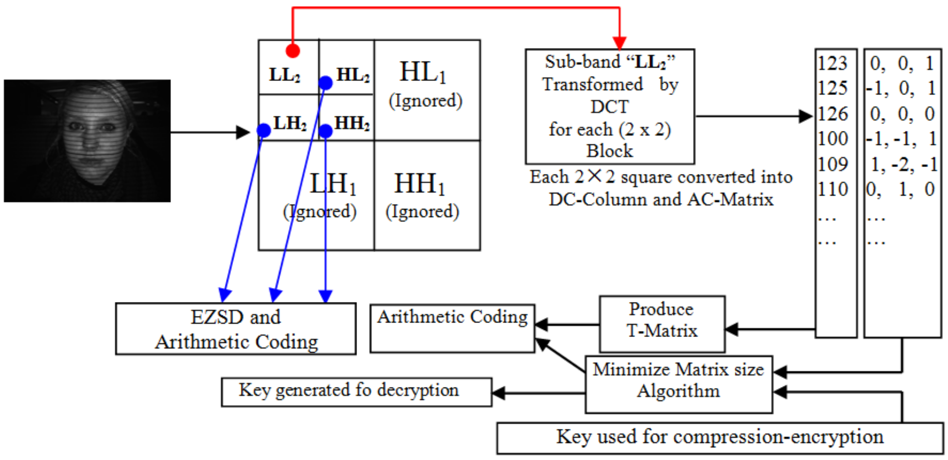


Figure 1. Layout of proposed Compression-Encryption algorithm

2.1 Discrete Wavelet Transform (DWT)

DWT analysis divides a signal into two classes (i.e. Approximation and Detail) by signal decomposition for various frequency bands and scales [1,2]. DWT utilizes two function sets: scaling and wavelet, which are associated with low and high-pass filters. In other words, only half of the samples in a signal are sufficient to represent the whole signal. The wavelet transform has some important properties; many of the coefficients for the high-frequency components (LH_1, HL_1 and HH_1) are zero or insignificant [5,6,10].This reflects the fact that much of the important information is contained in the LL_2 sub-band. In particular, the Daubechies wavelet transform has the ability to reconstruct approximately the original image by just using second level sub-bands (LL_2, HL_2, LH_2 and HH_2), while other sub-bands can be ignored. This property allows higher compression ratios[18,19].

2.2. Discrete Cosine Transform (DCT)

A second transform is appliedto each 2x2 block of pixels of LL_2 sub-band as show in Figure 2. The energy in the transformed coefficients is concentrated about the top-left corner of the matrix of coefficients. The top-left coefficients correspond to low frequencies: there is a 'peak' in energy in this area and the coefficients values rapidly decrease to the bottom right of the matrix, which means the higher-frequency coefficients [7,9]. The DCT coefficients are de-correlated, which means many of the small values (coefficients) can be discarded without significantly affecting image quality. A compact matrix of de-correlated coefficients can be compressed much more efficiently than a matrix of highly correlated pixels [8,17].

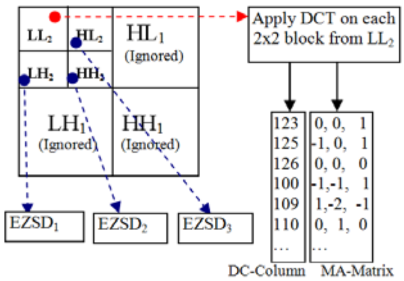


Figure 2. LL_2 sub-band quantized and transformed by DCT for each 2x2 blocks

Quantization is performed by matrix-dot-division and then truncating the result, by dividing each 2x2 coefficient from LL2. The quantized matrix removes insignificant coefficients. In the proposed method the high frequency sub-bands at first level are set to zero (i.e. discard HL_1, LH_1 and HH_1) as they do not affect image details. Additionally, only a small number of non-zero values are present in these sub-bands. In contrast, high-frequency sub-bands in the second level (HL_2, LH_2 and HH_2) cannot be discarded, as this would significantly affect image quality. For this reason, high-frequency values in this region are quantized. The quantization ***Q*** depends on the maximum value in each sub-band as follows:

$$Q = Quality * H_{\max} \tag{1}$$

Where the matrix $H_{\max}$ refers to the high-frequency coefficients in HL_2, LH_2 and HH_2, the factor Quality is used to increase/decrease. Thus, image details are reduced in case Quality >=0.01. The limit range for this factor is less than or equal to 0.9 for the 3D data used in this paper.

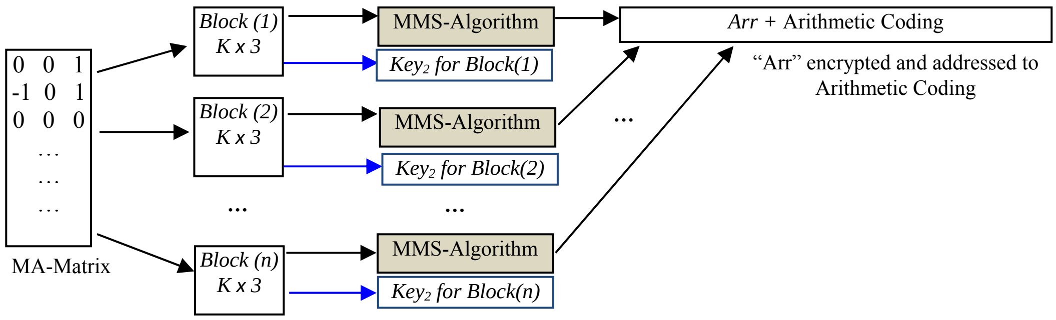
3. MINIMIZE-MATRIX-SIZE ALGORITHM (ENCRYPTION)

The purpose of this algorithm is to reduce the size of the MA-Matrix. This process depends on a key value and three adjacent coefficients to calculate and store the sum in a new array. The MMS algorithm consists of two parts: first, the MA-Matrix is partitioned into non-overlapping blocks (Kx3) of coefficients, where K refers to number of rows in a block as shown in Figure 3(a). Second, each block is encrypted by a *Key1*value. Additionally, new key values are generated for each block called *Key2*. Each key from each block is organized as minimum and maximum value for each column as shown in Figure 3(b). Each row and column *(r,c)*of the MA matrix is coded as follows:

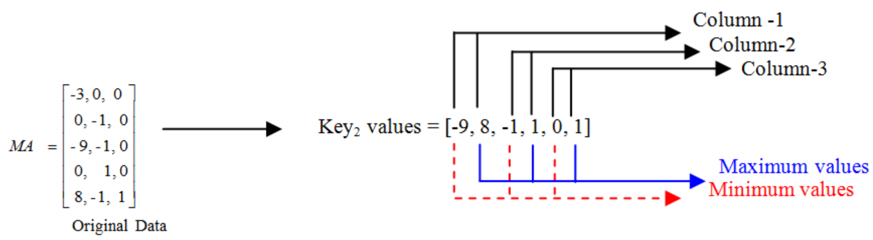
$$Arr_{(r,c)} = Key1_{(r,c)} * MA_{(r,c)} + Key1_{(r,c)} * MA_{(r,c)} + Key1_{(r,c)} * MA_{(r,c)} \tag{2}$$

ISBN: 978-989-8533-38-8 © 2015

The *Arr* contains stream of encrypted values. Thereafter, *Arr* is compressed by using arithmetic coding to produce stream of bits. The *key1* values are used in the Minimize-Matrix-Size Algorithm generated randomly, these key values are between {0…1} (for example; *Key1*= {0.128, 0.65, 0.8519}).



(a)



(b)

Figure 3. (a) MA-Matrix divided into blocks and each block encrypted by MMS-Algorithm, (b) Key_2 Values generated from a block, as an example block size =5.

4. TRANSFORMED MATRIX (T-MATRIX)

The DC-Column contains the DC values of DCT partitioned into 64-arrays. Each array is transformed by one-dimensional DCT, thereafter the quantization process is applied to each array as per Eq. (3) and then stored in a matrix called Transformed-Matrix (T-Matrix).

$$Q_n = Q_{(n-1)} + 1 \tag{3}$$

Where 64 ≥ n ≥1.The values in T-Matrix are de-correlated yielding good compression ratio. Each row of T-Matrix consists of low and high frequency coefficients. After scanning column-by-column, the T-Matrix is transformed into one-dimensional array which is then subject to Arithmetic Coding [8]. Figure 4 illustrates the process.

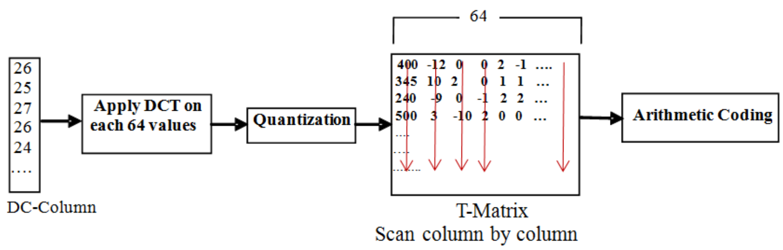


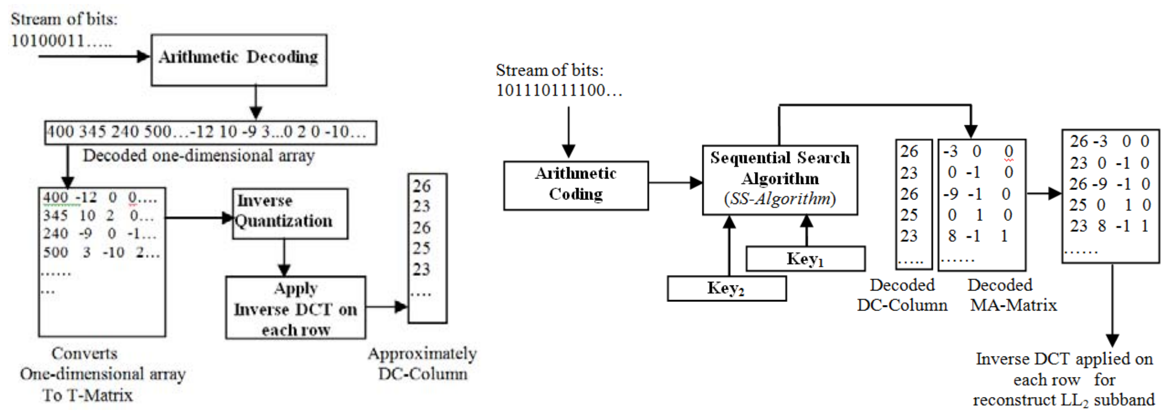
Figure 4. T-Matrix technique

5. ELIMINATE ZEROS AND STORE DATA (EZSD)

The EZSD algorithm is designed to increase compression ratio for high frequency sub-bands, and it is applied to each sub-band independently. It eliminates block of zeros, saving blocks of nonzero data in an array.The algorithm starts to partition the high-frequency sub-bands into non-overlapping 8x8 blocks, and then searches for nonzero blocks (i.e. search for at least one nonzero data inside a block). If the block contains any data, this block will be stored in the array called Reduced-Array; also the coordinates for the nonzero block are stored in new array called Positions. If the block contains just zeros, this block will be ignored, and the algorithm continues to search for other nonzero blocks. The final obtained Reduced Array is subject to Arithmetic Coding.

6. DECOMPRESSIONBY SEQUENTIAL SEARCH (SS-ALGORITHM)

The decompression-decryption algorithm consists of four steps. First, arithmetic decoding is used to recover the one-dimensional-array containing the original data in the T-Matrix, illustrated in Figure 5(a). Second, the novel SS-Algorithm is applied for decoding the MA-Matrix. This novel algorithm depends on the coded ***Arr***, Key_1 and Key_2 as illustrated in Figure 5(b). The encrypted array is partitioned into sub-arraysof size K. Each sub-array is subject to SS-Algorithm to recover the block of data in the MA-Matrix.



(a) (b)

Figure 5. (a) Decoding the DC-Column; (b) Decoding the MA-Matrix through the SS-Algorithm

The SS-Algorithm using three pointers, these pointers refer to original data in specific blocks of the MA-Matrix. The initial values of these pointers are set to minimum in the space search (Key2). These three pointers are called S1, S2 and S3 and are incremented by one in a gearwheel (e.g. similar to a clock, where S1, S2 and S3 represent hour, minutes and seconds respectively). At each iteration the SS-Algorithm computes the sum Eq. (4) and compares the error in Eq.(5)with zero. If true, the estimated values are S1, S2 and S3 corresponding to the original values in the MA-Matrix. In case of $E \neq 0$, the algorithm will continue to search for the original values in the block.

$$Sum = \sum_{i=1}^{3} S(i) \times Key_1(i) \qquad (4)$$

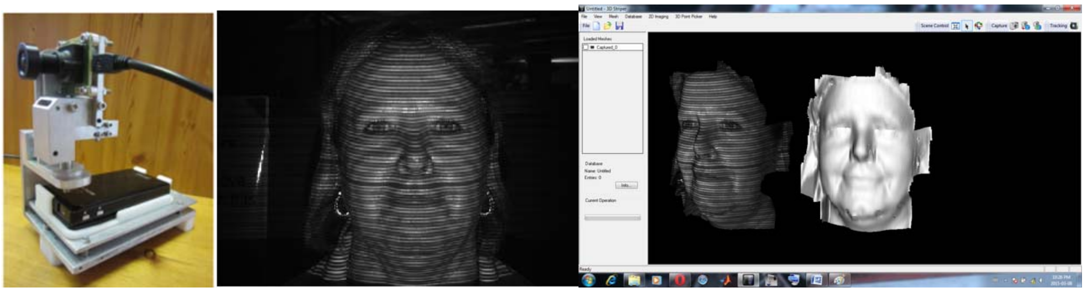
$$E = |Arr - Sum| \qquad (5)$$

At the third step, the decompression algorithm combines the DC-Column with MA-Matrix, and then using inverse DCT generates the LL2 sub-band. Next, the inverse EZSD recovers the high frequency sub-bands (LH2, HL2 and HH2); this algorithm recovers the locations and places the nonzero data in their exact locations in the recovered high-frequency matrix. Finally, the fourth step uses both inverse DWT and inverse DCT to decode 3D surface data recomposing all decompressed sub-bands.

ISBN: 978-989-8533-38-8 © 2015

7. EXPERIMENTAL RESULTS

The experimental results described here were implemented in MATLAB R2013a and Visual C++ 2008 running on an AMD Quad-Core microprocessor. We apply the compression and decompression algorithms to 2D images obtained from the GMPR scanner [13,14]; these images contain structured light patterns allowing 3D surface data to be generated from those patterns (Fig 6).The principle of operation of GMPR 3D surface scanning is to project patterns of light onto the target surface whose image is recorded by a camera. The shape of the captured pattern is combined with the spatial relationship between the light source and the camera, to determine the 3D position of the surface along the pattern. A surface can be scanned from a single 2D image and processed into 3D surface in a few milliseconds [15].



(a) (b)

Figure 6. (a)The 3D scanner captures a 2D image containing structured light**;** (b) 3D surface reconstructionfrom the 2D image using GMPR software

FACE1 FACE2 WALL

Figure 7. 2D images used in this research (dimension: 1392 x 1040 pixels, size:1.38 MB)

In this research we use 3 images depicted in Fig 7; these will be subject to compression followed by 2D to 3D reconstruction. The justification for introducing 3D reconstruction is that we can make use of a new set of metrics in terms of error measures and perceived quality of the 3D visualization to assess the quality of the compression and decompression algorithms. The rationale is that high quality image compression is required, otherwise the resulting 3D structure from the decompressed image will contain apparent dissimilarities and artefacts when compared to the 3D structure obtained from the original (uncompressed) data. We report on these differences through standard measures of RMSE-root mean square error as shown in Table 1. Additionally Fig. 8 shows visualization of 3D reconstruction compared with original (uncompressed) images.

Table 1.Compression-encryption applied to 2D images followed by 2D to 3D reconstruction

Image Name	Quality	Compressed Size (KB)	2D RMSE	3D RMSE
FACE1	0.01	49.6	5.19	2.0
	0.05	24.4	5.67	1.01
FACE2	0.01	47.8	4.42	4.23
	0.05	23.7	5.2	3.88
WALL	0.01	23.0	2.1	0.3
	0.05	11.63	2.66	1.4

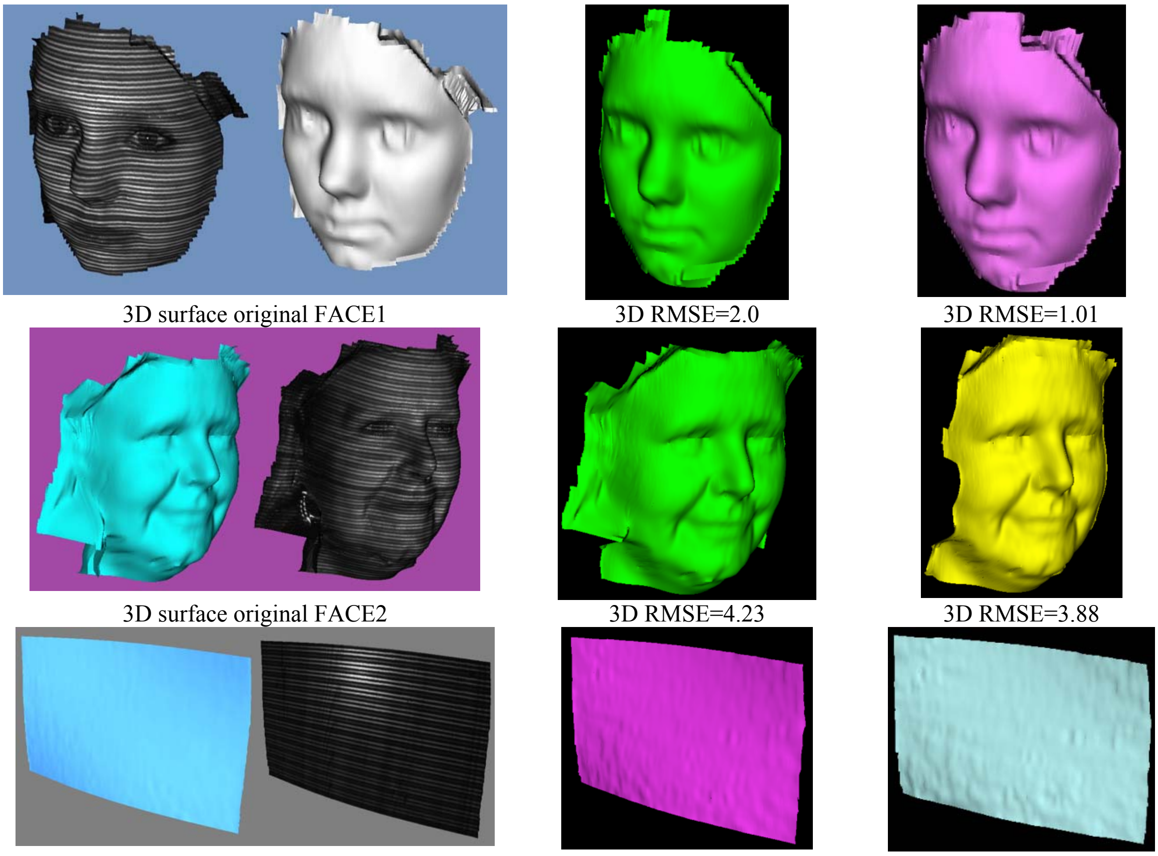


Figure 8. Comparison of 3D surface reconstruction using the original (uncompressed) images on the left, with compressed images on the right with RMSE measures.

8. CONCLUSIONS

This research has presented and demonstrated a novel method for 2D image compression and encryption. The quality of the method is illustrated through 2D to 3D reconstruction, 2D and 3D RMSE and the perceived quality of the visualization. The method is based on DWT and DCT in connection with the MMS algorithm. The results show that the proposed approach is capable of accurate 3D reconstructing with high compression ratios. The algorithm's advantages are highlighted as follows.

1. Using two transformations results in an increased number of high-frequency coefficients, leading to higher compression ratios.
2. The properties of the Daubechies wavelets are useful to obtain higher compression ratios; this is because high frequencies from the first level can be ignored without loss of accuracy.
3. The Minimize-Matrix-Size-Algorithm is used to partition the MA-Matrix into blocks and each row in each block are converted toa single value by Key_1. Additionally, it generates Key_2 for each block to increase the encryption level, thus Key_1 and Key_2 are both used in decryption step.
4. The SS-Algorithm (Decryption algorithm) represents the core of the decompression-decryption algorithm, which recovers the MA-Matrix by using encrypted data with Key_1 while $Key_2$is used to specify the space search for each block in the recovered MA-Matrix.
5. The EZSD algorithm used in this research removes block of zeros; at the same time it converts a high-frequency sub-band to an array containing few nonzero data, this process increases the compression ratio.

The approach disadvantages are summarized as follows. The overall complexity of the approach leads to increased execution time for both compression-encryption and decompression-decryption; the SS-Algorithm iterative method is particularly complex. Future work includes tackling the complexity of the method and developing alternative approaches to encoding and decoding the key management methods.

REFERENCES

[1] ADNAN KHASHMAN, KAMIL DIMILILER, 2008, Image Compression using Neural Networks and Haar Wavelet, *WSEAS TRANSACTIONS on SIGNAL PROCESSING*, Vol. 4, No.5.

[2] Ali Al-Haj, 2007, Combined DWT-DCT Digital Image Watermarking, *Science Publications, Journal of Computer Science* 3 (9): 740-746.

[3] Chen C. S. and Chen R. J, 2006, Image Encryption and Decryption using SCAN Methodology, *Proceedings of seventh International Conference on Parallel and Distributed Computing, Application and Technologies* (PDCAT'06), pp.61-66.

[4] El-Khamy S. E., Lotfy M. and Ali A. H, A New Color Image Encryption Technique utilizing Fuzzy Pseudo-random Bit Generator, *Proceedings of 22nd National Radio Science Conference* (NRSC), pp.185-194, Mar 2005.

[5] Grigorios D. , N. D. Zervas, N. Sklavos and Costas E. Goutis , 2008, Design Techniques and Implementation of Low Power High-Throughput Discrete Wavelet Transform Tilters for JPEG 2000 Standard, *WASET , International Journal of Signal Processing,* Vo. 4, No.1.

[6] G.Sadashivappa and K.V.S.Ananda Babu, 2008, PERFORMANCE ANALYSIS OF IMAGE CODING USINGWAVELETS, *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.10.

[7] I. E. G. Richardson, 2002, *Video Codec Design*, John Wiley & Sons.

[8] K.Sayood, 2000, *Introduction to Data Compression*, 2nd edition, Academic Press, Morgan Kaufman Publishers.

[9] K. R. Rao, P. Yip, 1990, *Discrete cosine transform: Algorithms, advantages, applications*, Academic Press, San Diego, CA.

[10] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, 1992, Image coding using wavelet transform," *IEEE Trans. on Image Processing*, Vol. 1, No. 2, pp. 205–220.

[11] Mitra A., Subba Rao Y. V. and Prasanna S. R. M, 2006, A New Image Encryption Approach using Combinational Permutation Techniques, *International Journal of Computer Science, Vol. 1, No. 2, pp. 127-131.*

[12] M. M. Siddeq, M. A. Rodrigues 2014a, A Novel Image Compression Algorithm for high-resolution 3D Reconstruction, *3D Research. Springer Vol. 5 No.2.*DOI 10.1007/s13319-014-0007-6

[13] M. M. Siddeq, M. A. Rodrigues, 2014b, A New 2D Image Compression Technique for 3D Surface Reconstruction, *18th International Conference on Circuits, Systems, Communications and Computers*, Santorin Island, Greece: 379-386

[14] M. Rodrigues, A. Robinson and A. Osman, 2010. Efficient 3D data compression through parameterization of free-form surface patches, In: *Signal Process and Multimedia Applications (SIGMAP), Proceedings of the 2010 International Conference on* IEEE, 130-135.

[15] M. Rodrigues, A. Osman and A. Robinson, 2013, Partial differential equations for 3D data compression and reconstruction, *Journal Advances in Dynamical Systems and Applications*, Vol. *8 No. 2, 303-315.*

[16] M. Rodrigues, M. Kormann, Schuhler and C. Tomek, P 2013. *Structured Light Techniques for 3D Surface Reconstruction in Robotic Tasks.* In: KACPRZYK, J, (ed.) Advances in Intelligent Systems and Computing.Heidelberg, Springer, 805-814.

[17] N. Ahmed, T. Natarajan and K. R. Rao, 1974.Discrete cosine transforms, *IEEE Transactions Computer*, Vol. C-23, pp. 90-93.

[18] Rafael C. Gonzalez, Richard E. Woods 2001, *Digital Image Processing*, Addison Wesley publishing company.

[19] T. Acharya and P. S. Tsai. 2005, *JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architecture*. New York: John Wiley & Sons.

[20] Tsai, M. and H. Hung, 2005, DCT and DWT based Image Watermarking Using Sub sampling, *in Proc. Of the 2005 IEEE Fourth Int. Conf. on Machine Learning and Cybernetics*, pp: 5308-5313, China.

[21] Yen J. C. and Guo J. I, 1999, A New Image Encryption System and its VLSI Architecture", *IEEE Workshop on Signal Processing Systems, Taipei, pp.430-437.*

[22] MM Siddeq and MA Rodrigues, 2015. A Novel 2D Image Compression Algorithm based on two level DWT and DCT Transforms with Enhanced Minimize-Matrix Size Algorithm for High Resolution Structured Light 3D Surface Reconstruction. Accepted by 3D Research (Springer).

FPGA BASED IMPLEMENTATION OF A NOVEL IMAGE STEGANOGRAPHY ALGORITHM

Mina Zolfy Lighvan , Masoom Nazari, Vahid Dinevari , Ziaeddin Daie Koozehkonani
and Ghader Karimian Khosroshahi
Department of Electrical & Computer Engineering, Tabriz University, Tabriz, Iran

ABSTRACT

As the complexity of current data flow systems and according infrastructure networks increases, the security of data transition through such platforms becomes more important. Thus, different areas of steganography turn to one of the most challengeable topics of current researches. In this paper a novel method is presented to hide an image into the host image and Hardware/Software design is proposed to implement our stagenography system on FPGA- DE2 70 Altera board. The size of the secret image is quadrant of the host image. Host image works as a cipher key to completely distort and encrypt the secret image using XOR operand. Each pixel of the secret image is composed of 8 bits (4 bit-pair) in which each bit-pair is distorted by XORing it with two LSB bits of the host image and putting the results in the location of two LSB bits of host image. The experimental results show the effectiveness of the proposed method compared to the most recently proposed algorithms by considering that the obtained information entropy for encrypt image is approximately equal to 8.

KEYWORDS: Host image, secret image, encryption, steganography, FPGA, DE2-70.

1. INTRODUCTION

The increasing complexity of data flow systems makes information security more important in data storage and transmission domains. Images are widely used in current processes and their protection from unauthorized access becomes more important. Fragile watermarking has been widely used to authentication and content integrity verification [1, 2]. Such a technique modifies the host image in order to insert the pattern but the permanent embedding distortion method is intolerable for the applications that require high quality images such as military images.

In addition, image encryption plays a significant role in information hiding [3-7]. Image hiding and encrypting algorithms range from simple spatial domain methods to more complicated and reliable frequency domains. Most of available encryption methods are proposed for common text data and are not proper for multimedia data such as images. Image encryption is somehow different from text encryption because of some inherent features of image, such as bulk data capacity and high correlation among its pixels. Thus, chaotic map ciphers and traditional cryptographic algorithms such as, RSA (Road Safety Authority) and DES (Data Encryption Standard) are no longer effective for practical image encryption, especially for an on-line communication scenario [8, 9]. In [10] image security is presented over internet through encryption/decryption of text data, however, due to large data size and real time constrains, algorithms that are good for textual data would not have the same performance on image.

Two levels of security are defined for digital image encryption as low level and high-level. In low-level security encryption, the encrypted image has degraded visual quality compared to that of the original one, but the content of the image is still visible and understandable to the viewers while in the high-level security, the content is completely scrambled and the image just looks like random noise.

In the most of natural images the values of the neighboring pixels are strongly correlated. This means that the value of any given pixel can be reasonably predicted from the values of its neighbors. Image encryption techniques try to convert an image to another one that is hard to detect. In our proposed method, the key idea is using both advantages of image watermarking and encryption in which the host image has been used both as a cipher key to distort the secret image and as host image to hide it. The hidden distorted image can easily be decrypted just by using the original host image. This technique allows transmitting high security images by hiding it in the ordinary image in which, because of a strong encryption and decreased correlation between encrypted image pixels, no one can doubt about that and predict it.

2. PROPOSED ALGHORITM

In this method both advantages of image encryption and watermarking is used for image steganography in which the host image is used as a cipher key for encryption and hiding the encrypted image.

2.1 Image Encryption Algorithm

Proposed method is based on distributing the secret image pixels into the host image pixels where, the size of host image is four times larger than the secret image. Using XOR operation, we can completely distort the secret image that won't be distinguishable. The most significant factor of this algorithm is hiding encrypted image inside the ordinary image that no one doubt about that and only can be decrypted by someone who has the original image. The schematic of the proposed algorithm for encryption is shown in Figure 1.

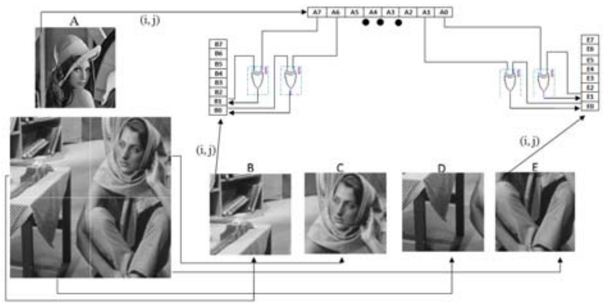


Figure 1. Block diagram of encryption system

Each pixel of secret image presented by 8 bits ($A_{7\text{-}0}$) is distributed through the pixels of the four different regions of host image. Each pixel of aforementioned regions takes two bits of the corresponding pixel of secret image (B, C, D and E take first, second, third and fourth two bits of the secret image, respectively) and XOR it with its own two LSB bits, then swap the result by its two LSB bits. This sequence is done for all pixels and as a result, each pixel of secret image is distributed in four pixels of host image. By changing of only two LSB bits of each pixel, the difference between original and watermarked image is so insignificant and can be neglected visually. As experimental results show, two LSB bits of each pixel in the host image have low cohesion with neighbor pixels. Thus to catch the best histogram and the best distorted image we used two LSB bits of pixels instead of other pairs. Using the mentioned method, an image has been hidden easily in the host image without making any visually significant change on it. The combination of pixels for encryption is detailed in Table 1.

Table 1. Combination of pixels for encryption

Main image	Host image	Watermarked image	Encrypted image
$A_{7\text{-}0}$	Region B : ($B_{7\text{-}0}$)	$B'_{7\text{-}0}$ = [($B_{7\text{-}2}$), ($B_{0\text{-}1} \oplus A_{6\text{-}7}$)]	$X_{7\text{-}0}$ = [($B_{0\text{-}1} \oplus A_{6\text{-}7}$),
	Region C : ($C_{7\text{-}0}$)	$C'_{7\text{-}0}$ = [($C_{7\text{-}2}$), ($C_{0\text{-}1} \oplus A_{4\text{-}5}$)]	($C_{0\text{-}1} \oplus A_{4\text{-}5}$),
	Region D : ($D_{7\text{-}0}$)	$D'_{7\text{-}0}$ = [($D_{7\text{-}2}$), ($D_{0\text{-}1} \oplus A_{2\text{-}3}$)]	($D_{0\text{-}1} \oplus A_{2\text{-}3}$),
	Region E : ($E_{7\text{-}0}$)	$E'_{7\text{-}0}$ = [($E_{7\text{-}2}$), ($E_{0\text{-}1} \oplus A_{0\text{-}1}$)]	($E_{0\text{-}1} \oplus A_{1\text{-}0}$)]

2.2 Image Decryption Algorithm

In order to extract the hidden image, a reversed sequence of encryption process is performed. Two LSB bits of each pixel of watermarked image in different regions mixed with the corresponding pixel in original host image using XOR operator and rearranged to create the secret image as shown in Figure 2.

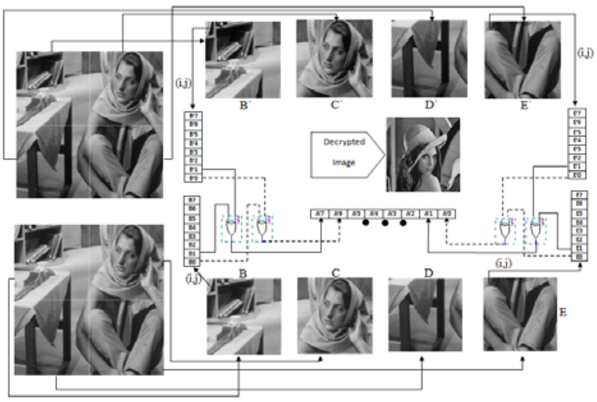


Figure 2. Block diagram of decryption system

Using the mentioned procedure, the extracted image is exactly same as the original image without any distortion. The combination and the order of pixels in decryption process are listed in Table 2.

Tabel 2. Combination of pixels for decryption

Host image	Watermarked image	Decrypted image
Region B : ($B_{7\text{-}0}$)	Region B' : ($B'_{7\text{-}0}$)	$A'_{7\text{-}0}$ = [($B_{1\text{-}0} \oplus B'_{1\text{-}0}$),
Region C : ($C_{7\text{-}0}$)	Region C' : ($C'_{7\text{-}0}$)	($C_{1\text{-}0} \oplus C'_{1\text{-}0}$),
Region D : ($D_{7\text{-}0}$)	Region D' : ($D'_{7\text{-}0}$)	($D_{1\text{-}0} \oplus D'_{1\text{-}0}$),
Region E : ($E_{7\text{-}0}$)	Region E' : ($E'_{7\text{-}0}$)	($E_{1\text{-}0} \oplus E'_{1\text{-}0}$)]

3. FPGA IMPLEMENTATION (ALTERA DE2-70)

Altera DE2-70 board is used for hardware implementing task which includes Touch-Panel LCD, Terasic TRDB_T5M (5 Megapixel camera), SD-Card, SSRAM and two 32MB SDRAM. The development is done using the Quartus II v.9 and NIOS II IDE. The implementation is divided into hardware and software sections. The image capturing is implemented in hardware, while image Steganography and displaying images in Touch-Panel LCD are done in software by the Nios II processor core. To have enough memory for frame buffering and also for manipulating data in Nios II, one of two 32MB SDRAM is used as frame buffer in hardware and other one is connected to Nios II (named as SDRAM0 and SDRAM1).

3.1 Hardware/Software Design

The hardware is created using the Hardware Description Language, Verilog and the c language based software accomplished the embedded system. The software is downloaded to NIOS II processor using NIOS II IDE which connected to CPU via JTAG UART. More details of Hardware/Software Design are explained as following.

3.1.1 Image Capture

The camera employed in this project is the Terasic TRDB-D5M [11]. To determine the size of captured frame, it is needed to configure the CMOS camera sensor. In operation, first a raw image is captured from the camera sensor, then translated to RGB (Red, Green, and Blue) format and temporarily stored in SDRAM0. The Verilog codes of these blocks are originally provided by Altera and only some slight changes applied to adapt them to the implementation requirements [12].

ISBN: 978-989-8533-38-8 © 2015

3.1.2 I2C Sensor Configuration

The Verilog HDL module that configures programming registers from the camera is I2C_CCD_Config.v. The aim of this module is to control and configure exposure time, resolution, and frame rate. This module uses the two-wire serial interface bus to communicate with TRDB-D5M registers. This HDL file was written originally by Altera. Only configuration values are changed to obtain output image with dimension of 800 × 480 as shown in Figure 3.

```
assign sensor_start_row        = 24'h010036;  // Start Row at 54
assign sensor_start_column     = 24'h020010;  // Start Column at 16
assign sensor_row_size         = 24'h0303BF;  // Row size is 960
assign sensor_column_size      = 24'h04063F;  // Column size is 1600
assign sensor_row_mode         = 24'h220011;  // Binning Row mode is 3
assign sensor_column_mode      = 24'h230011;  // Binning Column mode is 3
```

Figure 3. I2C Sensor Configuration values

The output image has a full resolution of 2592 × 1944, but by using Skipping and Binning, it can be reduced without affecting field-of-view. In skipping mode, entire rows and columns of pixels are not sampled. For example, 2X or 3X mode skips one or two "Bayer" pair of pixels, respectively, for every pair output. Binning reduces resolution by combining adjacent same-color pixels to produce one output pixel [12].

3.1.3 CMOS Sensor Data Capture

The Verilog HDL file that handles data capture from the camera is *CCD_Capture.v*. It is provided by Altera and no change is needed for this design. This module is responsible to get raw data from camera sensor and sends raw data with X and Y coordinates to next module.

3.1.4 Bayer Color Pattern Data to 30-bit RGB

The raw data from CMOS Sensor Data Capture module is divided into Red, Green and Blue colors by the module *RAW2RGB.v*. It is provided by Altera and no change is needed for this design.

3.1.5 Multi-Port SDRAM Controller

Captured frames are stored in the SDRAM0 temporarily. As mentioned earlier, Altera DE2-70 Development Board has two 32-MB synchronous DRAM. But only one of two 32 MB chips is used as frame buffer and the other one is added to NIOS II. To store frames, a FIFO (First In First Out) stack is used with dual-read and dual-write ports. Note that the read ports of this bank are connected to NIOS II via Avalon Bus and the write ports are connected to the output of RAW2RGB module. Thus, the bank ought to store at least 30 bits. But the ports have 16 bit Bus width. Thus a memory bank is proposed in which, the first 16 bits starting from address 0 stores 10 bit for red color and 5 bits for green color, and the second 16 bits starting from address 22'h700000 stores other 5 bits of green color plus 10 bits of blue color as shown in Figure 4.

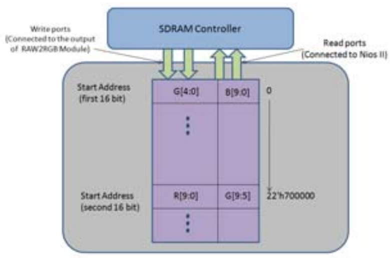


Figure 4. SDRAM0 memory bank

3.1.6 Nios II Processor Core and its Peripherals

Designing systems with embedded processors requires both hardware and software elements. Software involves the operations executed by processor (Nios II) and hardware includes operations directly executed by hardware on FPGA. Altera provides the SOPC Builder (System-On-a-Programmable Chip) that automates connecting software-hardware components to create a complete computer system. The soft-core processor used in this implementation is a Nios II processor. With SOPC Builder, it is possible to specify the settings

for a Nios II processor and add peripherals and select bus connections, I/O memory mappings, and IRQ assignments. Nios II core communicates with other modules using System Interconnect Fabric.
It is worth noting that Nios II/f core is selected for our system. It is optimized with dynamic branch prediction and 6 stage pipeline, 2Kbyte Data Cache and 4Kbyte Instruction Cache.

In the proposed implementation design, the host image is read from SD-card and the secret image is captured by CMOS slave controller. To establish a SD Card communication with Nios II, four PIO (named as sd_clk, sd_cmd, sd_dat, sd_dat3) are used that provide a synchronous serial data communication and The CMOS slave controller is originally developed by Altera. This module transfers temporarily stored frames in the SDRAM0 to the Nios II's processor, and also send a start/stop signal to CMOS sensor data capture module. The Verilog HDL of the CMOS Controller module is *cmos0.v*. To display the captured (secret image), host and the result image in the Touch panel LCD, LTM controller is added to Avalon Bus. This module uses SSRAM1 (second 32MB SDRAM) as image buffer.

4. EXPERIMENTAL ANALYSIS

Experimental analysis of the presented algorithm has been done on several images. Barbara image of size 512 × 512 is used as the host image and Lena image of size 256 × 256 as the image that will be hidden are shown in Figure 5.a. Figure 5.b shows the watermarked image which contains encrypted image. As can be seen, the watermarked image is visually same as the original host image. To see the result of encryption scheme, distorted image is extracted from the watermarked image and is compared with original secret image as shown in Figure 5.c.

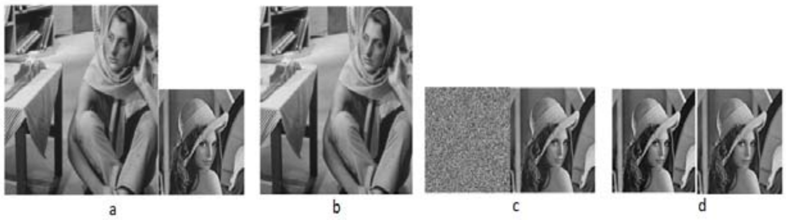


Figure 5. a) Barbara as host image and Lena as secret image b) Watermarked image with encrypted secret image c) Encrypted image and original image d) Left image is decrypted image and the right one is original secret image

The image extracted by the proposed decryption algorithm and the original image is shown in Figure 5.d. It can be seen that the decrypted image is absolutely clear and correct without any distortion.

Experimental result of XORing other bit-pairs of the host image for distorting the secret image are shown in Figure 6.Using the most significant bits of the host image to be XORed with the secret image bits, the distorted image is not good enough. MSB bits of neighbor pixels can't change the secret image as well as LSB bits because of high cohesion between MSB bits of image pixels. Due to the very low variations of significant bits of host image and as moving toward the least significant bit, the variations of bits increase in the image. Thus, only two LSB bits of each pixel are used to distort the secret image.

Figure 6. Results of encrypted image by using three other bit-pairs of host image: left, middle and right image is the result of respectively 7-6, 5-4 and 3-2 bits

4.1 Resistance to Statistical Attack

4.1.1 The Gray Histogram Analysis

To have an encrypted image with high-level of security, the histogram of the distorted image should be distributed uniformly between different levels of gray-scales. As Figure 7 shows, the best result is obtained using two least significant bits (as shown in Figure 7.d) and the reason of this uniformly distributed histogram lies on the fact that there is lower cohesion between LSB bits of image than MSB bits.

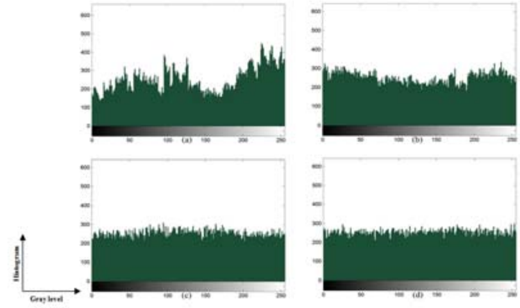


Figure 7. Histogram of encrypted image by using different bit-pairs of host image: (a) 7-6 bits (b) 5-4 bits (c) 3-2 bits (d) 1-2 bits.

With a statistical analysis of Lena image and its encrypted image, the corresponding greyscale histograms are extracted as shown in Figure 8. It demonstrates that the encryption algorithm has covered up all the characters of the secret image and shows good performance of balanced 0–1 ratio, zero co-correlation and high-level security. Comparing their histograms, we can find that the pixel grayscale values of the original image are concentrated on some values, but the histograms of the encrypted images are relatively uniform, which makes the statistical attacks difficult.

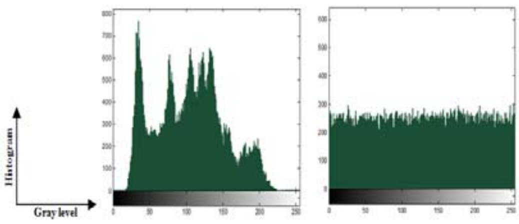


Figure 8. Left image is the histogram of Lena image and right image is its corresponding encrypted image using our proposed method

4.1.2 Correlation Coefficient Analysis

To test the correlation between two adjacent pixels in secret-image and ciphered image, the following procedure was carried out. First, randomly 1000 pairs of two adjacent (in horizontal, vertical, and diagonal direction) pixels from an image are selected. Then, the correlation coefficient of each pair is calculated by using Eq. (1) [13]:

$$r_{xy} = \frac{\mathrm{cov(x,y)}}{\sqrt{\mathrm{D(x)}}\sqrt{D(y)}} \tag{1}$$

$$\mathrm{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))(y_i - E(y)),$$

$$E(x) = \frac{1}{N}\sum_{i=1}^{N}x_i, \qquad D(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))^2$$

x and y are grey-scale values of two adjacent pixels in the image.

Figure 9 shows the correlation distribution of two vertically adjacent pixels in the secret-image and in the ciphered image. And the correlation coefficients are shown in Table 3. These correlation analysis prove that this algorithm satisfy zero co-correlation.

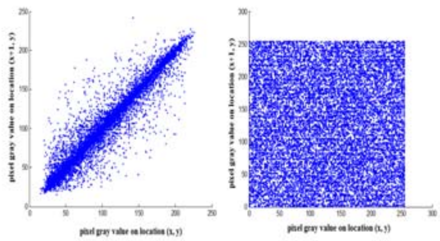


Figure 9. Correlations of two vertically adjacent pixels in the secret-image and in the cipher-image: left and right images are the correlation analysis of secret image and cipher-image

Tabel 3. Comparison between the correlation coefficients of Lena

	Horizontal	Vertical	Diagonal
Plain image(secret image)	0.9445	0.9701	0.9224
Encrypted image	-0.0024	0.0027	-.0036

4.1.3 Information Entropy Analysis

Information entropy is the most powerful feature that shows the randomness of image. Assume *m* as the information source, and the formula for calculating information entropy is as Eq. (2) [14]:

$$H(m) = \sum_{i=0}^{2^n-1} p(m_i)\log_2\left(\frac{1}{p(m_i)}\right) \tag{2}$$

Where *n* is the number of bits to represent a symbol $m_i \in m$. $p(m_i)$ Represents the probability of symbol m_i. For a purely random source emitting 2^n symbols,the entropy is $H(m) = n$. For encrypted images, the entropy should ideally be $H(m) = n$, for our work (8 bits) it is the information entropy of different image samples approximately equal to 8 and these results prove that the encrypted image is very hard to be predicted as shown in Table 4.

Tabel 4. Results of information entropy

	Lena	Airplane	Pepper
secret image	7.4273	6.7279	7.5748
Encrypted image(Ref. [3])	7.9874	7.9780	7.9860
Encrypted image(Proposed method)	**7.9974**	**7.9973**	**7.9972**

4.2 Similarity Measurements between Host and Watermarked Image

To compute the similarity between the host image and watermarked image, three different similarity measures is considered: Mean Square error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Correlation. These are given by Eq. (3), Eq. (4) and Eq. (5) [15].

$$MSE = \left(\frac{1}{X \times Y}\right)\sum_{x=1}^{X}\sum_{y=1}^{Y}(I(x,y) - G(x,y))^2 \tag{3}$$

Where *MSE is* Mean Square Error. X and Y are the dimensions of the image. I and G are the original host and the watermarked image respectively.

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \tag{4}$$

Where *PSNR* is Peak Signal-to-Noise Ratio.

$$r_{IG} = \frac{\sum_{x=1}^{X}\sum_{y=1}^{Y} I(x,y) \times G(x,y)}{\sum_{x=1}^{X}\sum_{y=1}^{Y} G(x,y)^2} \tag{5}$$

Where r_{IG} is the correlation between I and G.
The results for Similarity measurements are shown in Table 5.

Table 5. Similarity measurements between original host image and watermarked image

MSE	*PSNR*	r_{IG}
.7275×10^{-4}	180.1791	0.9926

5. CONCLUSION

In this paper a robust steganography technique has proposed for image security transmission and a Hardware/Software co-design scheme is presented to implement our stagenography algorithm on FPGA-DE2-70 board. The aim is to distribute each pixel of secret image between four pixels of host image. Due to using only two LSB bits of each pixel value for hiding image, there is insignificant difference between host image and watermarked image. The similarity of them is very critical for high security transmission. Experimental results show that, the proposed algorithm outperforms than the most recently presented technique. Lowest correlation between secret image and encrypted image, uniform histogram of encrypted image and the highest information entropy are three main features of the proposed method.

REFERENCES

[1] H. Guo, Y. Li, A. Liu, and S. Jajodia, "A fragile watermarking scheme for detecting malicious modifications of database relations," *Information Sciences,* vol. 176, pp. 1350-1378, 2006.

[2] P. MeenakshiDevi, M. Venkatesan, and K. Duraiswamy, "A fragile watermarking scheme for image authentication with tamper localization using integer wavelet transform," *Journal of Computer Science,* vol. 5, p. 831, 2009.

[3] M. A. Singh, M. L. Goswami, and A. U. R. Khan, "A Novel Image ENCRYTION Algorithm for Enhanced Security," in *International Journal of Engineering Research and Technology*, 2013.

[4] S. Kumar, "Image Steganography Using Improved LSB And Exor Encryption Algorithm," THAPAR UNIVERSITY, 2014.

[5] S. Hui, "Comparison and Analytical Research on Several Digital Image Watermark Encryption Algorithms," *Journal of Beijing Institute of Graphic Communication,* vol. 4, p. 014, 2013.

[6] X. Huang and G. Ye, "An image encryption algorithm based on hyper-chaos and DNA sequence," *Multimedia Tools and Applications*, pp. 1-14, 2012.

[7] Z. Pengtao, L. Gang, W. Minghao, and P. Jieli, "An image encryption algorithm based on multi-scroll chaotic map," in *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, 2012, pp. 186-189.

[8] G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos, Solitons & Fractals,* vol. 21, pp. 749-761, 2004.

[9] M. J. Wiener, "Cryptanalysis of short RSA secret exponents," *Information Theory, IEEE Transactions on,* vol. 36, pp. 553-558, 1990.

[10] D. S. A. Minaam, H. M. Abdual-Kader, and M. M. Hadhoud, "Evaluating the Effects of Symmetric Cryptography Algorithms on Power Consumption for Different Data Types," *IJ Network Security,* vol. 11, pp. 78-87, 2010.

[11] M. Petouris, A. Kalantzopoulos, and E. Zigouris, "An FPGA-based Digital Camera System Controlled from an LCD Touch Panel," in *Signals, Circuits and Systems, 2009. ISSCS 2009. International Symposium on*, 2009, pp. 1-4.

[12] F. Filipe, "IP Camera on FPGA with a Web Server," Phd, Minho, 2010.

[13] X. Tong and M. Cui, "Image encryption with compound chaotic sequence cipher shifting dynamically," *Image and Vision Computing,* vol. 26, pp. 843-850, 2008.

[14] C. Rey and J.-L. Dugelay, "A survey of watermarking algorithms for image authentication," *EURASIP Journal on Applied Signal Processing,* vol. 2002, pp. 613-621, 2002.

[15] S. Li, X. Mou, and Y. Cai, "Improving security of a chaotic encryption approach," *Physics Letters A,* vol. 290, pp. 127-133, 2001.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

STATIC AND DYNAMIC OBJECT RECONSTRUCTION USING PHASE SHIFTING OF DE BRUIJN PATTERN

Matea Đonlić, Tomislav Petković, and Tomislav Pribanić
University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia

ABSTRACT

A novel structured light method for color 3D surface profilometry is proposed. The proposed method may be used for reconstruction of both static and moving objects and does not require accurate color calibration of a camera–projector pair. The method is based on a structured light pattern that is a combination of a De Bruijn color sequence and of a sinusoidal fringe pattern. For static scenes phase-shifting and the De Bruijn window property are combined to obtain a high accuracy reconstruction. For dynamic scenes the Hessian ridge detector is applied to find stripes whose colors are decoded using a Gaussian mixture model. The De Bruijn window property and dynamic programming then uniquely identify projected stripes yielding a spatially sparser reconstruction. We have tested the proposed method on multiple objects with challenging surfaces and different albedos that clearly demonstrate both usability and robustness.

KEYWORDS

3D surface profilometry, pattern projection, colored structured light, phase-shifting, De Bruijn sequence

1. INTRODUCTION

Structured light projection is a widely studied topic in the field of 3D imaging where reconstruction needs to be fast, noninvasive and inexpensive. One of the main challenges is improving the accuracy of the reconstruction while projecting as few patterns as possible. That goal is important because reduced acquisition time enables the reconstruction of moving objects or of objects that cannot stand still, e.g. humans. Non-contact 3D surface profilometry based on structured light is the method of choice in reverse engineering, industrial quality control, object recognition, health-care applications, and others.

Based on the scene type structured light methods may be classified into dynamic and static scene reconstruction methods. For dynamic scenes only one pattern is projected, so these methods are also called one-shot methods. For static scenes multiple structured light patterns are projected over time, so these are called multiple-shot methods. Multiple-shot methods produce superior reconstruction than single-shot methods in terms of both resolution and accuracy, but have a substantially greater acquisition times and are mostly unsuitable for dynamic scenes. There are approaches proposing the use of multiple images for dynamic scene reconstruction but expensive high-speed hardware is required (Sagawa et al., 2014; Narasimhan et al., 2008; Zhang and Huang, 2006). A more detailed classification of various reconstruction methods and structured light patterns is given in Salvi et al. (2010).

Several widely used one-shot methods (Zhou et al., 2012; Zhang et al., 2002) are based on color stripe patterns that satisfy a window property. A finite sequence of length L satisfies the window property if every subsequence of length N, called window, appears exactly once. The window property enables extraction of spatial position through identification of unique subsequences in a color sequence. Usual problems all such approaches must solve are robust and precise detection of color stripe centers and accurate color decoding that is not affected by transfer functions of projector-camera pair and by object albedo.

A large number of state-of-the art multiple-shot methods are based on multiple phase-shifting or on a combination of Gray or color coding and phase-shifting (Salvi et al., 2010). In phase-shifting a set of sinusoidal patterns is projected over time and pixel codification is based only on the recorded values in time. The main difference between various methods is how the problem of phase unwrapping is solved. Phase estimation, a core of all such methods, returns a phase estimate that is ambiguous by an unknown multiple of 2π and is therefore called wrapped-phase. The wrapped phase must be robustly unwrapped to reconstruct a

3D surface. This is achieved by projecting multiple spatial frequencies from which a required offset value may be computed; or by using additional images that contain Gray code (Sansoni et al., 2003) or a color-coded pattern (Gorthi and Lolla, 2005) that uniquely identifies the required multiple of 2π. Some authors propose combining different patterns in multiple channels of a color image (Pan et al., 2006; Wust and Capson, 1991) effectively reducing the number of images and shortening the acquisition time; such approaches give accurate unwrapped phase but cannot cope with discontinuities on objects with complicated surfaces (Zhang, 2012).

Additional problems arise when color is used: different projector-camera transfer functions for color channels, object albedo, and ambient lightning. Many authors tried to solve these problems by using different variants of color calibration to eliminate undesired effects. Such approaches usually involve projection of multiple solid color patterns from which the crosstalk matrix, the albedo coefficients, and the ambient light illumination are estimated (Zhou et al., 2012; Zhang et al., 2002; Caspi et al., 1998). Alternatively, a tedious photometric calibration of a camera-projector pair may be performed (Juang and Majumder, 2007).

Both De Bruijn sequences and sinusoidal fringes are often independently used in 3D surface profilometry. In this paper we propose a novel approach to 3D reconstruction of both dynamic and static scenes based on a color structured light pattern that combines a De Bruijn color sequence and a sinusoidal fringe. Compared to the current methods described in the literature the proposed method does not require precise color calibration of the camera-projector pair as: (1) for one-shot reconstruction the colors are classified and recognized using Gaussian mixture models (GMM), and (2) for multiple-shot reconstruction acquired images are self-corrected for albedo and ambient light by using the specific property of the proposed structured light pattern. There has been some previous work in various fields on the topic of image segmentation using GMM with images in RGB (Wu et al., 2003), HSV (Huang and Liu, 2007), or HSI (Henderson et al., 2008) color spaces but, to the best of our knowledge, GMM have not been applied to 3D surface profilometry.

2. PROPOSED METHOD

The proposed method uses a novel structured light pattern that enables both coarse one-shot reconstruction and fine multiple-shot reconstruction. In the following sections we first present the proposed structured light pattern and describe how it is constructed. Then we give descriptions of the proposed one-shot reconstruction for dynamic scenes and of the proposed multiple-shot reconstruction for static scenes.

2.1 Structured Light Pattern

Proposed structured light pattern is comprised of multiple color images represented in three-channel RGB color space that are produced by combining a De Bruijn color sequence and a sinusoidal fringe pattern.

A De Bruijn sequence of order N over an alphabet of size K is a pseudo-random cyclic sequence of length K^N (MacWilliams and Sloane, 1976) in which every possible subsequence of length N (window) appears exactly once. Color coding in every image is based on a De Bruijn sequence of order $N = 3$ over an alphabet of $K = 6$ colors. Used colors are red, yellow, green, cyan, blue, and magenta (Table 1). We also impose two additional constraints on the sequence: (1) any two adjacent colors in the generated sequence must be different, i.e. RR, GG, YY, BB, MM, and CC are prohibited; and (2) in every subsequence of length N both minimum and maximum in all three channels must be achieved at least once, e.g. BMC is not allowed as all values in blue channel are 1 (at least one 0 is required), similarly RGB is allowed as both 0 and 1 appear at least once in all channels. The first condition improves the robustness of color decoding for dynamic scenes. The second condition is required for static scene reconstruction as it enables pixel-based color equalization that effectively eliminates object albedo and enables robust temporal color decoding. Therefore, the proposed sequence satisfying the second condition is effectively albedo self-correcting.

De Bruijn sequences are generated as Hamiltonian paths in a De Bruijn graph of order N where all graph vertices that do not satisfy the aforementioned two conditions are removed. The maximal length of a sequence for $N = 3$ and $K = 6$ that satisfies both conditions is = 90. An example sequence is:

$$S = \left\{ \begin{matrix} \mathrm{RYBRGCRGBRCRCYRCGRCBYRBYGBYCMRGMRCMYGMYBYBGRB} \\ \mathrm{GYBCRBCYBMGRMGYMGCMGMCRMCYMCGMBYMBGMGBMYCBRYC} \end{matrix} \right\}. \tag{1}$$

The intensity of a sinusoidal fringe pattern is defined as

$$I(x) = I_{\max}\left(0.5 - 0.5\cos\left(\tfrac{2\pi}{P}x - \varphi_i\right)\right), \tag{2}$$

where I_{max} is the maximum intensity, x is image column index, P is the period, and φ_i is phase shift.

Table 1. An alphabet of six colors from which a De Bruijn sequence is constructed

Color	Letter	(b,g,r)	Hue	Color	Letter	(b,g,r)	Hue
Red	**R**	(0,0,1)	0°	Yellow	**Y**	(0,1,1)	60°
Green	**G**	(0,1,0)	120°	Cyan	**C**	(1,1,0)	180°
Blue	**B**	(1,0,0)	240°	Magenta	**M**	(1,0,1)	300°

To obtain one image in the proposed structured light pattern we combine the color sequence (1) and the fringe pattern (2) in the HSV color space. The sequence (1) defines the hue channel: one hue value is used per one period of the fringe pattern, therefore there are exactly L colored fringes in the final image. Saturation is always set to 1. The fringe pattern (2) defines the value channel. The pattern is comprised of $N \cdot N_p$ phase shifted images, where N is window length and N_p is the number of phase shifts per one fringe period. For the ith image in the pattern the sequence S is rotated by $mod(i+N_p-1,\ N_p)$ symbols and $I(x)$ is shifted by $\varphi_i = i\frac{2\pi}{N_p}$, where $i = 0, 1, \ldots, N \cdot N_p - 1$. An example of the proposed pattern is shown in Figure 1; note how the entire pattern shifts to the right and how used colors wrap around from the end to the beginning of the sequence.

The proposed pattern may be classified as a multi-slit pattern as the intensity modulation given by (2) has a side effect of introducing a black slit between every two color stripes. Every image from the pattern may be used for one-shot reconstruction of dynamic scenes, and $N \cdot N_p$ images may be used for multiple-shot reconstruction of static scenes, i.e. *the same pattern is used for both one and multiple-shot reconstruction.*

$i = 0,\ \varphi_i = 0$

$i = 4,\ \varphi_i = 2\pi$

$i = 8,\ \varphi_i = 4\pi$

Figure 1. Three of twelve images from the proposed structured light pattern for L= 90 and N_p = 4

2.2 One-Shot Reconstruction

One-shot reconstruction is comprised of the following three steps: (a) detection of stripes and theirs centers in the acquired frame, (b) identification of stripe color, and (c) spatial decoding of the De Bruijn code.

2.2.1 Stripe Segmentation

Colors stripes in the image are detected using the multi-scale vesselness measure that is based on a 2D ridge detector that uses a Hessian matrix. Detection consists of the following steps: (a) input image is converted from RGB to HSV color space and V channel is extracted, (b) the vesselness map is computed as described in (Frangi et al., 1998) at scales matching the stripe width, (c) centers of stripes are extracted as positions of the local maxima in the vesselness map, and (d) sub-pixel positions of centerlines are obtained as described in (Steger, 1998). Parameters used for computation of the vesselness map are automatically selected as described in (Petković and Lončarić, 2010).

2.2.2 Color Identification and Stripe Positioning

We propose using standard clustering techniques in the RGB color space for color identification. Clustering will enable separation of observed colors into the required number of distinct groups and will thus mitigate the difficult color matching problem between projector and camera that is usually solved by a tedious color calibration procedure. A convenient fact that boosts the effectiveness of clustering techniques is *a priori* knowledge of the number of colors that are projected. This prior information about number of colors enables use of the Gaussian mixture model (GMM) and Expectation-maximization (EM) algorithm to separate the observed colors into six groups. A similar approach was used by Fechteler and Eisert (2008), where authors fitted the pixel data to prototype color lines inside of the RGB cube.

To use GMM and EM the number of components in the mixture must be known and each component must be described with a Gaussian (normal) distribution defined by its mean and covariance matrix. Once the data is recorded a posterior probability for each component may be computed and the component parameters (mean and covariance matrix) may be adjusted to better fit the data. These steps are then repeated until convergence. For a more detailed explanation of GMM and EM see Bishop (2006).

Our proposed structured light pattern uses six colors (Table 1) therefore GMM must have six mixtures. Processing will be done in RGB color space as used colors are vertices of a RGB cube. We set the means of the six color components in the middle of color axes in the RGB cube. The covariance matrices of each component are chosen in the shape of an elongated ellipsoid oriented in the direction of its color axis. As the result of EM algorithm we assign each stripe center to a corresponding color cluster based on the resulting probability that the center belongs to that cluster. In this way a color is assigned to each stripe in the image.

After assigning colors to the stripes we compare each row in the image to the projected De Bruijn color sequence. From each image row we extract color labels and apply dynamic programming to find the best correspondence between the two color sequences. For finding the best correspondences we use a similar approach as one proposed by Zhang et al. (2002).

2.3 Multiple-Shot Reconstruction

Multiple-shot reconstruction is based on the phase-shifting paradigm. Both phase and color subsequence that are required for the 3D reconstruction are derived using temporal analysis for each pixel independently, therefore there are no assumptions about spatial characteristics of the imaged object. Proposed reconstruction is done in the following three steps: (a) channel equalization and albedo removal, (b) estimation of the wrapped phase, and (c) temporal color decoding and phase unwrapping.

2.3.1 Temporal Channel Equalization

For every pixel and for every channel let I_{min} and I_{max} be observed minimal and maximal values for the recorded $N \cdot N_p$ frames. Then each of RGB channels is separately equalized via affine mapping

$$I_{eq}(i) = \tfrac{255}{I_{\max} - I_{\min}} \left(I(i) - I_{\min} \right),\ i = 0, \ldots, N \cdot N_p - 1. \tag{3}$$

This temporal channel equalization effectively removes the albedo for static pixels and is supported by the property we imposed to the De Bruijn sequence: in every subsequence of length N both minimum and maximum in all three channels must be achieved at least once; observed values I_{min} and I_{max} are these extrema.

2.3.2 Estimation of Wrapped Phase

For every pixel the wrapped phase is estimated from the V channel of HSV image representation as

$$\psi_w = a\tan 2\left(-\textstyle\sum_{i=0}^{N \times N_p - 1} V_{eq}(i)\sin(\varphi_i), \sum_{i=0}^{N \times N_p - 1} V_{eq}(i)\cos(\varphi_i) \right) \tag{4}$$

where i is the frame index. Note that the V channel of HSV representation must be computed using the equalized channel data of Eq. (3) as $V_{eq}(i) = \max\left(B_{eq}(i), G_{eq}(i), R_{eq}(i) \right)$, otherwise the result will be affected by the object albedo.

2.3.3 Temporal Color Decoding and Phase Unwrapping

To unwrap the wrapped phase ψ_w we must identify a De Bruijn subsequence of length $N = 3$ for every pixel independently. Combining Eqs. (2) and (4) for a fixed pixel yields a three parameter temporal intensity model for the V channel of HSV representation,

$$I(i) = I_0 + I_1 \cos(\psi_w + \varphi_i), \tag{5}$$

where I_0, I_1, and ψ_w are model parameters. The ith recorded frame has phase $\varphi_i = \frac{2\pi}{N_p} i$ so temporal intensity given by Eq. (5) achieves maxima at $\psi_w + \varphi_i = 2k\pi$. To identify the length N subsequence we must use N sample points at maximal phases $\varphi_i = -\psi_w + 2k\pi,\ k \in Z,$ which correspond to real values $i_R = -\frac{\psi_w}{2\pi} N_p + kN_p,\ k \in Z,$ that fall into [0, $N \cdot N_p$–1] interval. By taking two frames with indices i closest to the three chosen values of i_R and by linearly interpolating between them we obtain three RGB triplets that correspond to the three maxima in V channel. Color is identified by computing the hue and selecting the closest color. Time-reversed color sequence then identifies the phase offset needed for unwrapping, e.g. if a time-reversed triplet starts at position m in the sequence given by Eq. (1), then the phase offset is $2\pi(m+1)$ and the unwrapped phase is $-\psi_w + 2\pi(m+1)$. An example of the temporal color decoding for N_p = 4 and P = 11 is shown in Fig. **2**. A function given by Eq. (5) fitted to the V channel is shown as a solid black line. Three sample points where colors are decoded are marked with vertical black lines. Decoded color substring is YCR. Its time-reversed value RCY is found at the position m = 12 so phase offset is $2\pi(12+1)$.

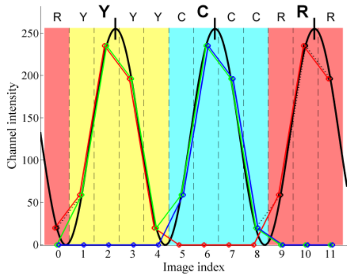


Figure 2. Example of the temporal color decoding on simulated data.
Phase offset is given by the position of RCY subsequence in the used De Bruijn sequence.

3. RESULTS AND DISCUSSION

All presented experiments were performed using Acer X1260 DLP projector (resolution 1024×768) that was paired with PointGrey DragonFly2 DR-HICOL camera (resolution 1024×768) fitted with Fujinon HF9-HA1B lens. Geometric calibration of the camera (and projector) was performed using the procedure described in Zhang (2000). Color calibration was not performed, i.e. factory settings were used. Camera was white-balanced for all acquisitions. The camera-projector distance was about 11cm and the working distance was about 1m, giving the calibration volume of approximately $500 \times 500 \times 500$ mm. We evaluated our methods by fitting 3D planes to the reconstructed planar surfaces. The average residual error of the fitted 3D planes is 0.7459 ± 0.5917 mm for the multiple-shot and 0.9795 ± 0.749 mm for the one-shot method.

3.1 One-Shot Reconstruction

We have proposed a GMM clustering in the RGB space for color identification as the hue component alone cannot be used for robust color identification due to significant color shifting caused by object's albedo. This is clearly demonstrated in Figure 3 for a human face: compare the histogram of hue values and the scattergram of RGB points and note that a clear separation between six colors disappears if only hue is considered.

An important fact in using GMM for color identification is initial positioning of clusters. If the number of data points of one color is significantly different than from other colors then it may be necessary to adjust initial parameters. This problem may be mitigated by ensuring proper sequence properties: in any area of the image colors on stripes must be equally distributed. Also, initial centers may be adjusted by using equalization parameters I_{min} and I_{max} of the multiple-shot method as they provide sufficient information about object albedo for every pixel.

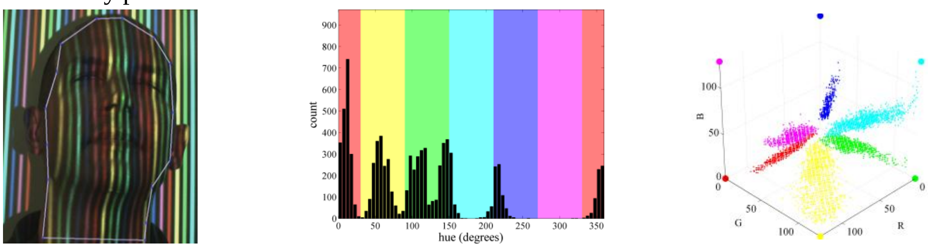


Figure 3. Left: manually segmented human face. Middle: a histogram of hue channel. Right: a scattergram of RGB data.
Note significant color shifts and unclear boundaries for red and magenta in the hue histogram

ISBN: 978-989-8533-38-8 © 2015

We have tested our method on various objects, including artificial material and human skin, to assess robustness to different albedos. Fig. 4 shows that using the proposed method both mannequin surface and human skin are accurately reconstructed. Reconstructions may be incomplete due to occlusions, shadows or dark parts of objects since in those parts we cannot detect stripe centers.

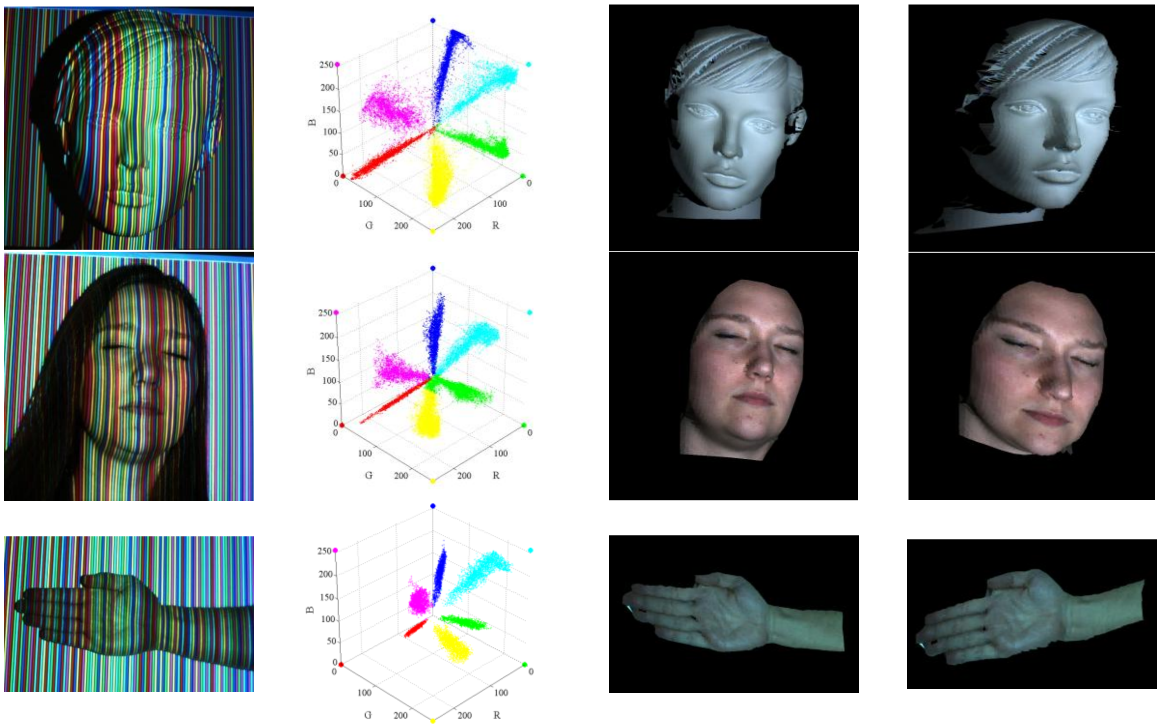


Figure 4. Examples of 3D reconstructions for one-shot method for three different objects.
First column holds input images, second color distributions, and third and fourth textured 3D reconstructions

3.2 Multiple-Shot Reconstruction

Multiple-shot reconstruction may be used for all pixels that are static. For those pixels the method uses the proposed second property that makes color calibration unnecessary. This is demonstrated in Figure 5: note how the equalized image has significantly clearer colors than the unprocessed camera image; dynamic range in black areas of the tripod was about 10/255 and the colors were still successfully extracted.

Reconstruction of a human face using one- and multiple-shot method on the same data is shown in Figure 6. Note how the multiple-shot method gives higher spatial resolution than the one-shot method. To assess the accuracy of the proposed multiple-shot reconstruction method we have compared wrapped phases for two recorded objects, a cube on a tripod (Figure 5) and a human face (Figure 6), to ones obtained using traditional greylevel sinusoidal fringe phase shifting. Histograms of phase errors are shown in Figure 7.

We have demonstrated that the proposed methods accurately reconstruct both artificial objects with uniform albedo and human skin, a fairly challenging surface. One-shot method provides a reconstruction that is limited by the numbers of stripes that can be projected, as expected. This shortcoming is mitigated by a multiple-shot method that enables dense 3D surface reconstruction for static objects.

4. CONCLUSION

The main challenge in colored structured light is the correct matching of colors perceived by the camera to the colors projected by the projector. We have proposed two techniques to solve this problem: using clustering techniques in the one-shot and using the color correcting sequence in the multiple-shot method.

We have demonstrated the proposed method’s usability on various objects with challenging surfaces and with different albedos. Methods successfully identify colors without the need for color calibration even for objects such as human face where colors are strongly modified with the skin color absorption and reflectance. The one-shot reconstruction gives a lower spatial resolution than the multiple-shot in horizontal direction. Depth resolution is preserved due to a Hessian ridge detector that gives subpixel-precise stripe positions.

Our future efforts are directed toward creating a system for automatic detection of static and dynamic objects in the scene and applying the proposed one- and multiple-shot methods where appropriate in an approach similar to one proposed by Zhang et al. (2014). Creating such system would enable a full usage of our proposed pattern for the accurate reconstruction of both static and dynamic objects at the same time.

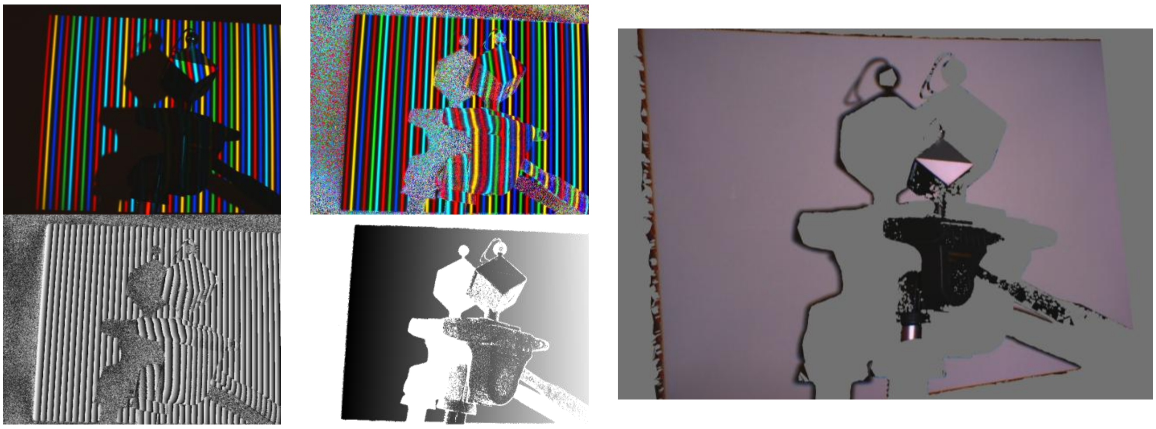


Figure 5. The SpyderCUBE™ on a tripod. Top-left is the input image (1/12), top-middle is equalized image (1/12), bottom-left is wrapped phase, bottom-middle is unwrapped phase, and right is 3D reconstruction

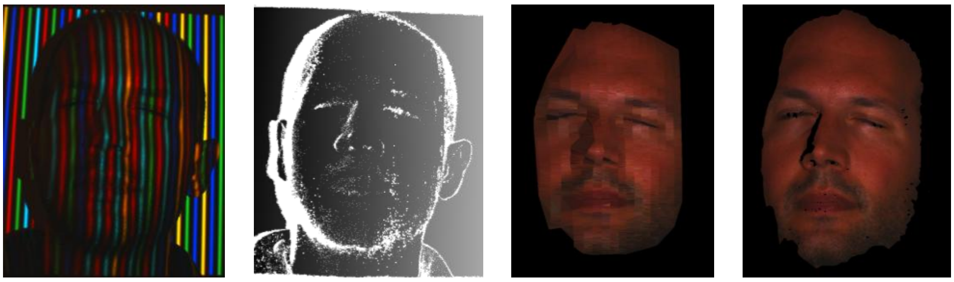


Figure 6. The human face. From left to right: input image, unwrapped phase, textured 3D reconstruction using one-shot method, and textured 3D reconstruction using multiple-shot method

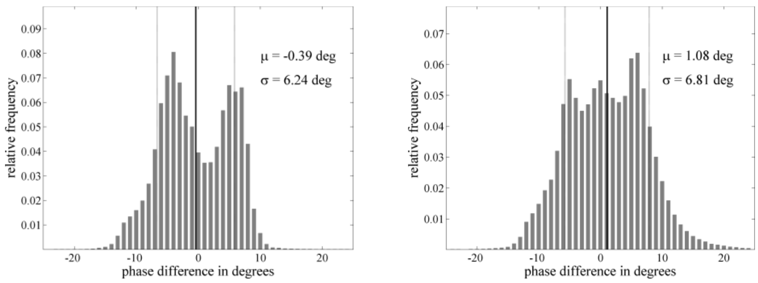


Figure 7. Histograms of phase errors. Left: a cube on a tripod from Fig. 5. Right: a human face from Fig. 6

ACKNOWLEDGEMENT

This work has been supported in part by Croatian Science Foundation's funding of the project IP-11-2013-3717. We are grateful to Joaquim Salvi, Professor of Computer Vision at University of Girona, Spain, who suggested improvements to the content and presentation of this paper.

REFERENCES

Bishop, C., 2006. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, USA.

Caspi D. et al., 1998. Range Imaging With Adaptive Color Structured Light. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 20, No. 5, pp. 470-480

Fechteler, P. and Eisert, P., 2008. Adaptive Color Classification for Structured Light Systems. In *IEEE Comput. Soc. Conf. on Comput. Vision and Pattern Recognition Workshops*. Anchorage, Alaska, USA, pp. 1-7

Frangi A.F. et al., 1998. Multiscale vessel enhancement filtering. In W.M. Wells, A.Colchester, and S.Delp (eds.), *Medical Image Computing and Computer-Assisted Interventation,* Springer Berlin Heidelberg, pp.130-137.

Gorthi, S.S. and Lolla, K.R., 2005. New Approach for Simple and Rapid Shape Measurement of Objects with Surface Discontinuities. In *Optical Meterology*, *Proc. of SPIE*, Vol. 5856, pp. 184-194.

Henderson N. et al., 2008. An Application of Gaussian Mixtures: Colour Segmenting for the Four Legged League Using HSI Colour Space. In U. Visser. et al. (eds.), *RoboCup 2007: Robot Soccer World Cup XI*, Vol. 5001 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 254-261.

Huang, Z.K. and Liu D.H., 2007. Segmentation of Color Image Using EM algorithm in HSV Color Space. In *Int. Conf. on Inform.Acquisition*. Seogwipo, South Korea, pp. 316-319.

Juang, R, and Majumder, A., 2007. Photometric Self-Calibration of a Projector-Camera System. In *IEEE Conf. on Comput.Vision and Pattern Recognition*. Minneapolis, Minnesota, USA, pp. 1-8.

MacWilliams, F.J. and Sloane, N.J.A., 1976. Pseudo Random Sequences and Arrays. In *Proc. of the IEEE*, Vol. 44, No. 12, pp. 1715-1729

Narasimhan S.G. et al., 2008. Temporal Dithering of Illumination for Fast Active Vision. In D. Forsyth and P. Torr, and A. Zisserman (eds.), *European Conf. on Comput. Vision;* Vol. 5305 of *Lecture notes in Comput. Sci.*, Springer Berlin Heidelberg, pp. 830-844.

Pan J. et al., 2006. Color Phase-shifting Technique for Three-dimensional Shape Measurement. *Optical Engineering*, Vol. 45, No. 1, pp. 013602-013602-9.

Petković, T. and Lončarić, S., 2010. Using X-ray imaging model to improve guidewire detection. In *IEEE 10th Int. Conf. on Signal Process.,* Beijing, pp. 805-808.

Sagawa, R. et al., 2014. Dense 3D Reconstruction from High Frame-Rate Video Using a Static Grid Pattern. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 36, No. 9, pp. 1733-1747.

Salvi J. et al., 2010. A State of the Art in Structured Light Patterns for Surface Profilometry. *Pattern Recognition*, Vol. 43, No. 8, pp. 2666-2680

Sansoni G. et al., 2003. OPL-3D: A novel, portable optical digitizer for fast acquisition of free-form surfaces. *Review of Scientific Instruments,* Vol. 74, No. 4, pp. 2593-2603.

Steger, C., 1998. An Unbiased Detector of Curvilinear Structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 20, No. 2, pp.113-125.

Wu Y. et al., 2003. Unsupervised Color Image Segmentation Based on Gaussian Mixture Model. In *Proc. of the Joint Conf. of the 4th Int. Conf. on Inform., Commun. and Signal Process. and 4th Pacific Rim Conf. on Multimedia, Vol. 1*. Singapore, pp. 541-544.

Wust, C. and Capson, D.W., 1991. Surface Profile Measurement Using Color Fringe Projection. *Mach. Vision and Applicat.*, Vol. 4, No. 3, pp. 193-203

Zhang L. et al., 2002. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. In *Proc. of the 1st Int. Symp. on 3D Data Process. Visualization and Transmission*. Padova, Italy, pp. 24-36

Zhang, S. and Huang, P.S., 2006. High-resolution, Real-time Three-dimensional Shape Measurement. *Optical Engineering*, Vol. 45, No. 12, pp. 123601-1 - 123601-8.

Zhang Y. et al., 2014. Real-Time Scalable Depth Sensing With Hybrid Structured Light Illumination. *IEEE Trans. Image Process.*, Vol. 23, No. 1, pp. 97-109.

Zhang, Z., 2000. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, No. 11, pp. 1330-1334.

Zhang, Z.H., 2012. Review of single-shot 3D shape measurement by phase calculation-based fringe projection techniques. *Optics and Lasers in Engineering*, Vol. 50, No. 8, pp. 1097-1106.

Zhou Y. et al., 2012. Adaptive Color Calibration Based One-Shot Structured Light System. *Sensors*, Vol. 12. No. 8, pp. 10947-10963.

COMPARING THE PERFORMANCE OF RECOVERY ALGORITHMS FOR ROBUST FACE RECOGNITION

Sedigheh Ghofrani, Seyedramin Alikiaamiri and Mehran Khorasani
Electrical and Electronic Engineering Department, Islamic Azad University, South Tehran Branch, Tehran, Iran.

ABSTRACT

The problem of automatically recognizing human faces from the facial images under varying expression and illumination is still attractive. The sparse signal representation is a new topic for object classification and face recognition. Implementing the sparse representation needs using a proper reconstruction algorithm to recover the sparse signal as well. In this paper, the performance of three greedy algorithms named matching pursuit (MP), orthogonal matching pursuit (OMP), compressive sampling matching pursuit (CoSaMP) and two ℓ_1-minimization algorithms named primal-dual interior-point (PDIP) and Homotopy are compared according to achieved accuracy, consuming time, and the number of iterations. For this purpose, three well known datasets named, AR, Extended Yale B and Essex University are used.

KEYWORDS

Face Recognition, Sparse Representation, Recovery Algorithm.

1. INTRODUCTION

The well-known Shannon-Nyquist sampling principal [1] indicates that the sample rate frequency must be at least twice the signal maximum frequency for the perfect reconstruction. Compressive sensing (CS) theory introduced in 2006 [2] is a novel approach for compressible signal sampling and reconstructing in a significantly lower sample rate than Shannon-Nyquist. In general, CS uses non-adaptive linear projections that preserve the structure of a signal and reconstruct the signal by using an optimization process on projections [2]. Generally, CS needs two fundamental conditions, sparsity and incoherency. The term incoherency, is defined as a pair of orthogonal bases which are used is CS algorithm, so that this pair satisfies incoherent sampling condition [2]. Many natural signals are sparse when expressed by convenient basis. Through this concept, several signals such as audio and image are sparse when represented in appropriate basis domain.

With the rapidly increasing demand on face recognition technology, it is not surprising to see an overwhelming amount of research publications like CS for this topic in recent years [3]. Researchers used CS to constitute appropriate low-dimensional dictionary and approve results of the face recognition. Because of lower sample rate that is needed in CS theory, some features such as robustness to pose and expression can also be reachable [4]. Two common procedures for any applications based on CS theory are sensing or measuring a signal and representing or recovering the signal [5].

Several reconstruction or recovery algorithms in CS based on ℓ_1-minimization have been proposed. Among them, greedy algorithms due to fast implementation and low complexity of mathematical framework are well known [6]. The performances of greedy algorithms are evaluated by the reconstruction error and signal to noise ratio. In general, the main challenges of choosing an efficient algorithm for under-determined problems are being more sparse and fast. In addition, various reconstruction algorithms may be suitable for different problems. In this paper, we use matching pursuit [7] (MP), orthogonal matching pursuit [6] (OMP) and compressive sampling matching pursuit [6] (CoSaMP) as greedy algorithms and Homotopy [8] and primal-dual interior-point (PDIP) [9], [10] for signal reconstruction. The performances of these five mentioned reconstruction algorithms are evaluated for three different face datasets, named AR [11], Extended Yale B [12] and Essex University [13].

The paper is organized as follows. In Section 2, the face recognition based on sparse representation approach is explained. In Section 3, the reconstruction algorithms for signal recovery are reviewed. Experimental results are shown in Section 4, and finally we conclude this work in Section 5.

2. FACE RECOGNITION BASED ON SPARSE REPRESENTATION

A pioneer paper implementing sparse representation for face recognition was published in 2009 [14]. According to [14], images of *i*-th class are concatenated into $\mathbf{A_i} = [\mathbf{p_{i1}},\ldots,\mathbf{p_{in_i}}]$, where $\mathbf{p_{in_i}}$ with size $M\times 1$ refers to the *n*-th image of *i*-th class. Any new probe image $\mathbf{y}$ associated with *i*-th class is expressed as,

$$\mathbf{y} = \alpha_{i_1}\mathbf{p_{i_1}} + \alpha_{i_2}\mathbf{p_{i_2}} + \ldots + \alpha_{i_n}\mathbf{p_{i_{n_i}}} , \tag{1}$$

where simply can be written in matrix form,

$$\mathbf{y} = \mathbf{A_i}\boldsymbol{\alpha}_i \tag{2}$$

where $\boldsymbol{\alpha}_i = [\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_n}]^T$ is a vector with size $n\times 1$ including coefficients and $(.)^T$ refers the transpose operator. Considering $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A_i}, \ldots, \mathbf{A}_C]$ with size $M\times N$ including the all training samples for all C classes where $N = nC$, Eq. (2) is rewritten as,

$$\mathbf{y} = \mathbf{Ax} \tag{3}$$

where $\mathbf{x} = [0,\ldots,0, \alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_n}, 0,\ldots,0]^T$ with size $N\times 1$. Obviously the only nonzero coefficients of $\mathbf{x}$ belong to the *i*-th class, therefore $\mathbf{x}$ is *n*-sparse. In general $\mathbf{A} \in \mathbf{R}^{M\times N}$ is a full rank matrix ($M << N$) named dictionary, $\mathbf{y} \in \mathbf{R}^M$ is the measurement vector or the probe image and $\mathbf{x} \in \mathbf{R}^N$ is a *n*-sparse vector. In practice, Eq. (3) is inherently under-determined, because the number of images (*N*) belong to the *C* classes are greater than the number of equations (*M*). So the probe image $\mathbf{y}$ can be classified by solving the ℓ_1-norm minimization as,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_{\ell_1} \quad \text{subject to} \quad \mathbf{y} = \mathbf{Ax} \tag{4}$$

or considering the optional error tolerance ε ,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_{\ell_1} \quad \text{subject to} \quad ||\mathbf{y} - \mathbf{Ax}||_{\ell_2} < \varepsilon \tag{5}$$

The procedures of recovery algorithms are explained in Section 3.

For identifying the probe image $\mathbf{y}$, the sparse coefficients $\hat{\mathbf{x}}$ should be obtained by solving Eq. (4) or alternatively Eq. (5). Ideally, nonzero entries in $\hat{\mathbf{x}}$ are to be associated with a specified class, however, practically, nonzero entries may also be associated with other classes. So, as said in [14], we obtain $\boldsymbol{\delta}_\mathbf{i}(\mathbf{x})$, that is a new vector which selects only nonzero coefficients of $\hat{\mathbf{x}}$ that are related to *i*-th class. Then the residual for each class is computed as

$$r_i(\mathbf{y}) = ||\mathbf{y} - \mathbf{A}\boldsymbol{\delta}_\mathbf{i}(\mathbf{x})||_{\ell_2} \tag{6}$$

where $i = 1,2,\ldots,C$. The probe image $\mathbf{y}$ is belonging to the class with minimum residual value. The method named SRC and the procedures are shown in Fig. 1.

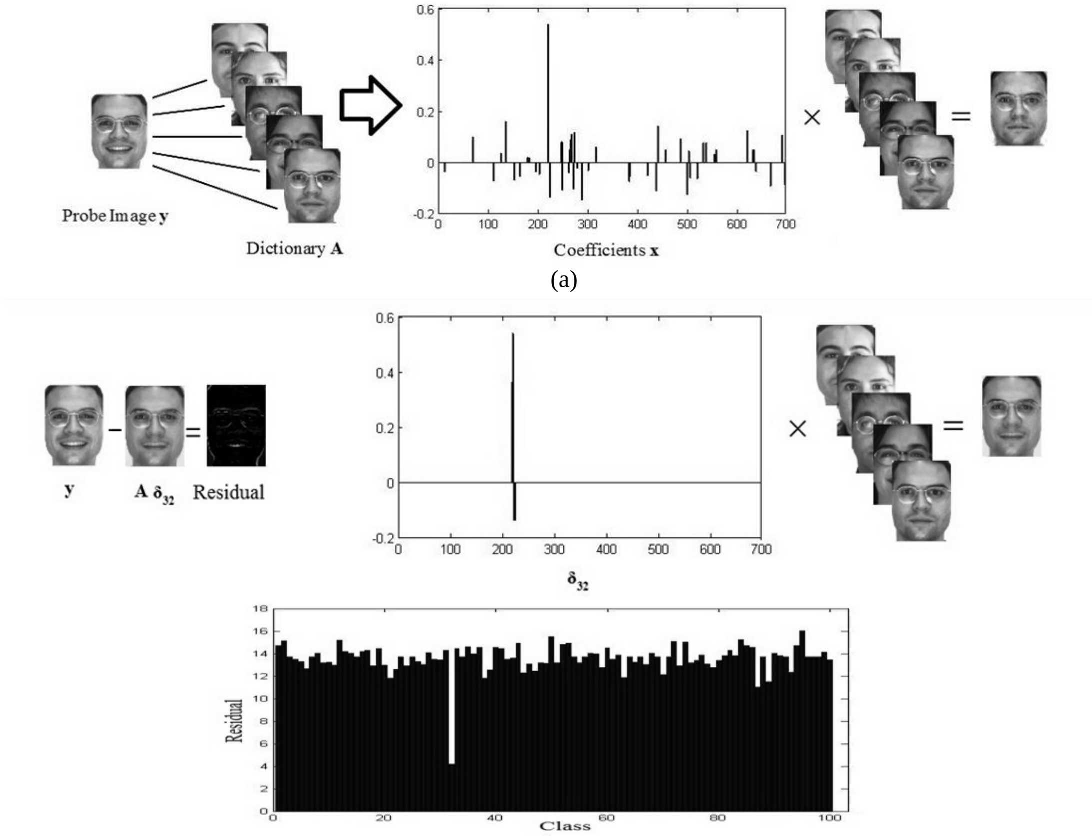


Figure.1. The SRC method is shown (a) obtaining the sparse coefficient vector $\mathbf{x}$ for a sample probe belongs to AR dataset, (b) computing the residuals for every class based on Eq. (6).

3. CS RECOVERY ALGORITHMS

In this work, five reconstruction algorithms are used to solve the ℓ_1-norm optimization problem according to Eq. (4) or alternatively Eq. (5). As we ourselves have implemented MP, OMP and CoSaMP, in following, they are explained in detail. For implementing the MP and OMP only one atom is chosen at each iteration, whereas $2K$ atoms are selected for CoSaMP at every iteration. In addition, the obtained coefficients are not changing for MP, and they are updated for OMP in every iteration. Although CoSaMP needs less iteration number in comparison with MP and OMP, the computation time of CoSaMP is considerable.

The first step is normalizing the dictionary $\mathbf{A}$ to have unit ℓ_2 norm, then, set $\mathbf{x_0} = 0 \in \mathbf{R}^N$ as initial vector value and $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x_0} \in \mathbf{R}^M$ as initial residual vector and determine the integer value of K as the sparsity level, for MP, OMP and CoSaMP algorithms. In addition, we set $\mathbf{\Lambda}_0 = \Phi$ for OMP and CoSaMP where Φ is a null matrix. The size of matrix $\mathbf{\Lambda}_j$ is growing while the number of iteration is increasing, and $j = 1,2,\ldots,J$ denotes the iteration number. The following stages are performed after the above mentioned setting for every iteration, $j = 1,2,\ldots,J$.

ISBN: 978-989-8533-38-8 © 2015

- **MP**

1. Compute the inner product $\mathbf{g}_j = \mathbf{A}^{\mathbf{T}}\mathbf{r}_{j-1} \in \mathbf{R}^N$.
2. Find the index k where: $k = \arg\max_{i=1,\ldots,N} |\mathbf{g}_j[i]|$.
3. Update $\mathbf{x}_j[k] = \mathbf{x}_{j-1}[k] + \mathbf{g}_j[k]$ where $\mathbf{x}_j[k]$ and $\mathbf{g}_j[k]$ refer to the k-th element in $\mathbf{x}_j \in R^N$ and $\mathbf{g}_j \in R^N$. Then, compute the residual $\mathbf{r}_j = \mathbf{r}_{j-1} - \mathbf{g}_j[k]\mathbf{A}_k$ where $\mathbf{A}_k$ is the k-th column of dictionary $\mathbf{A}$.
4. Consider the iteration number, j, or obtain the residual norm value, $||\mathbf{r}_j||_{\ell_2}$. Stop the algorithm if j is greater or $||\mathbf{r}_j||_{\ell_2}$ is less than the pre-defined value. Otherwise go through the first step.

The two iterations of MP for a sample probe image is shown in Fig. 2.

Figure 2. The procedure of MP algorithm for a sample image chosen from AR dataset. (Top) The first iteration, (Middle) the second iteration, (Bottom) the recovered image and computed coefficients.

- **OMP**

1. Compute the inner product $\mathbf{g}_j = \mathbf{A}^{\mathbf{T}}\mathbf{r}_{j-1} \in \mathbf{R}^N$.
2. Find the index k where $k = \arg\max_{i=1,\ldots,N} |\mathbf{g}_j[i]|$.
3. Update the index set and matrix of chosen atoms $\boldsymbol{\Lambda}_j = \boldsymbol{\Lambda}_{j-1} \cup \{k\}$, $\mathbf{A}_{\Lambda_j} = \mathbf{A}_{\Lambda_{j-1}} \cup \mathbf{A}_k$.
4. Obtain the new estimate $\tilde{\mathbf{x}} = (\mathbf{A}_{\boldsymbol{\Lambda}_j}^{\mathbf{T}}\mathbf{A}_{\boldsymbol{\Lambda}_j})^{-1}\mathbf{A}_{\boldsymbol{\Lambda}_j}^{\mathbf{T}}\mathbf{y}$. Note that the size of $\tilde{\mathbf{x}}$ is growing while the number of iteration is increasing. Compute the coefficient vector $\mathbf{x}_j[\boldsymbol{\Lambda}_j] = \tilde{\mathbf{x}}$.
5. Update the residual $\mathbf{r}_j = \mathbf{y} - \mathbf{A}\mathbf{x}_j$.
6. Consider the iteration number, j, or obtain the residual norm value, $||\mathbf{r}_j||_{\ell_2}$ Stop the algorithm if j is greater or $||\mathbf{r}_j||_{\ell_2}$ is less than the pre-defined value. Otherwise go through the first step.

The two iterations of OMP for a sample probe image is shown in Fig. 3.

- **CoSaMP**

  The procedures of CoSaMP are exactly the same as OMP, except for every iteration, we choose the $2K$ largest index of k, where $k = \arg\max_{i=1,\dots,N} |\mathbf{g}_j[i]|$.

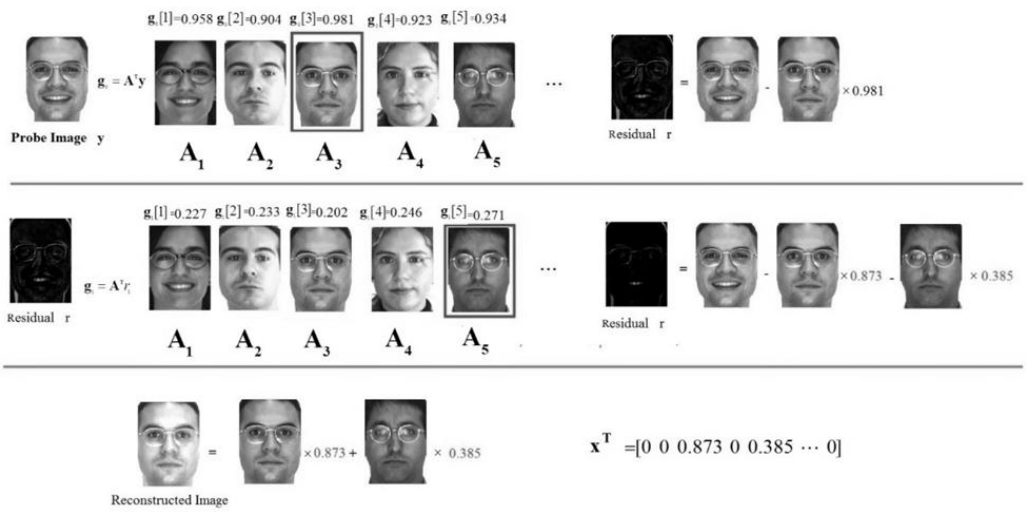


Figure 3. The procedure of OMP algorithm for a sample image chosen from AR dataset. (Top) The first iteration, (Middle) the second iteration, (Bottom) the recovered image and computed coefficients.

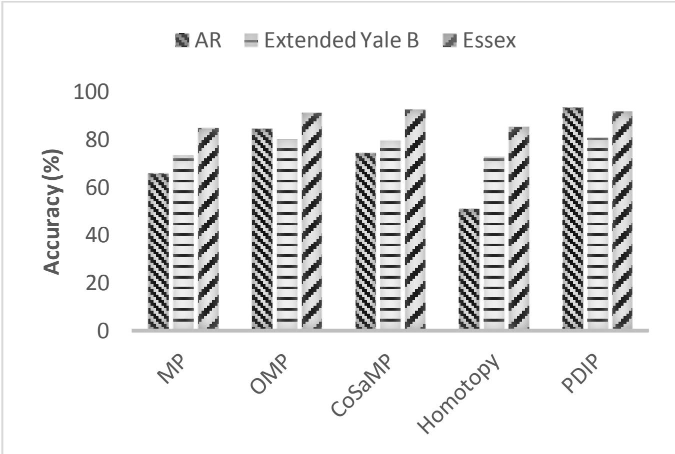
4. EXPERIMENTAL RESULTS

The main purpose of this paper is comparing the performance of various reconstruction algorithms under the same circumstances. So, three datasets named AR, Extended Yale B and Essex University are used. AR dataset includes more than 4000 frontal face images of 126 individual persons. From AR dataset, 50 male and 50 female subjects are chosen, and for every individual person, 14 images with varying illumination and expression are used. As images of AR were taken at two different sessions [15], 7 images of first session as training set and 7 images of second session as the test set are used. Extended Yale B includes 2414 frontal face images of 38 individual persons captured under different illumination conditions. For Extended Yale B, half of images for training set and the other half as test set are chosen randomly. Essex face dataset contains 3040 frontal images of 152 individuals. The images were taken in 20 frames when a person was talking. For Essex dataset, one image as training set and 19 images as test set are used for every class.

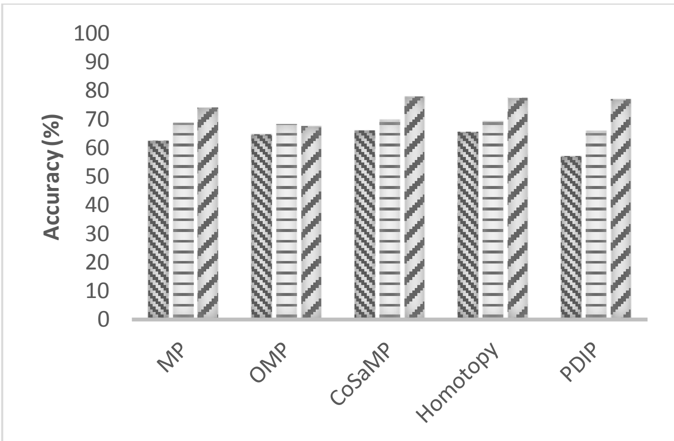
In this paper we formed two types dictionary by face images and LoG filtered face images [16]. As we have used SRC for face recognition, the method based on using the first dictionary named SRC and the method based on using the second dictionary named SRC-LoG. We exploit the discriminative nature of sparse representation to perform classification. If sufficient training samples are available for every class, the test samples can be represented as a linear combination of just those training samples from the same class. This representation is naturally sparse, involving only a small fraction of the overall training dataset. It is actually the sparsest linear representation of the test sample in terms of this dictionary and can be recovered efficiently via reconstruction algorithms [14]. By appropriate dictionary constitution, each person has his own class in the dictionary. If a proper recovery algorithm is used, the sparse response for a test image would be obtained. On the other hand, sparse response for the test image can choose the class of that image in a dictionary correctly. This means that the more sparse coefficients, the more reliable face recognition. As we said above, signal sparsity depends on efficient dictionary generation and appropriate reconstruction algorithm. As MP and OMP select one atom of the dictionary at each iteration, if an atom is chosen incorrectly at the first iteration, the error is spreading among all iterations. It means that the recovered image has no similarity to the probe input image. To overcome this problem, CoSaMP selects $2K$ atoms of the dictionary at every iteration and updates them in next step, so we expect that CoSaMP almost outperforms both MP and OMP. The experimental results shown in Fig. 4 approve the expectation. Moreover, another

ISBN: 978-989-8533-38-8 © 2015

important factor for comparing these algorithms is the running time. These algorithms are based on either error or iteration number. So, it takes time to recognize the probe face correctly. This factor is important because it affects the efficiency of face recognition algorithm. In Fig. 4, the results show that PDIP algorithm is more efficient than other algorithms for most datasets. However, the running time of the PDIP algorithm for total evaluation of each dataset is more than other reconstruction algorithms. The results of comparison between the iteration number (J) and the computation time for AR dataset are shown in Fig. 5. According to the experimental results, PDIP algorithm has the best performance for the most data sets and OMP algorithm is the fastest algorithm that can recognize faces in a short time. All algorithms are implemented using MATLAB on a PC with Core i7 CPU and 4GB of RAM.



(a)



(b)

Figure 4. Comparison the performance of five different CS recovery algorithm for different datasets when (a) SRC, (b) SRC-LoG are used.

5. CONCLUSION

Implementing the sparse representation for object classification needs an appropriate recovery algorithm. In this work, the performance of MP, OMP, CoSaMP, PDIP and Homotopy as the reconstruction algorithm for face recognition are evaluated when three well known datasets named AR, Extended Yale B and Essex University are used. According to the experimental results, all algorithms have their own drawbacks and advantages. However, OMP, CoSaMP and PDIP achieved an appropriate accuracy in comparison with MP and Homotopy. Furthermore, OMP and CoSaMP need less computation time than PDIP. At the end, the LoG operator is not improving the face recognition accuracy via SRC approach.

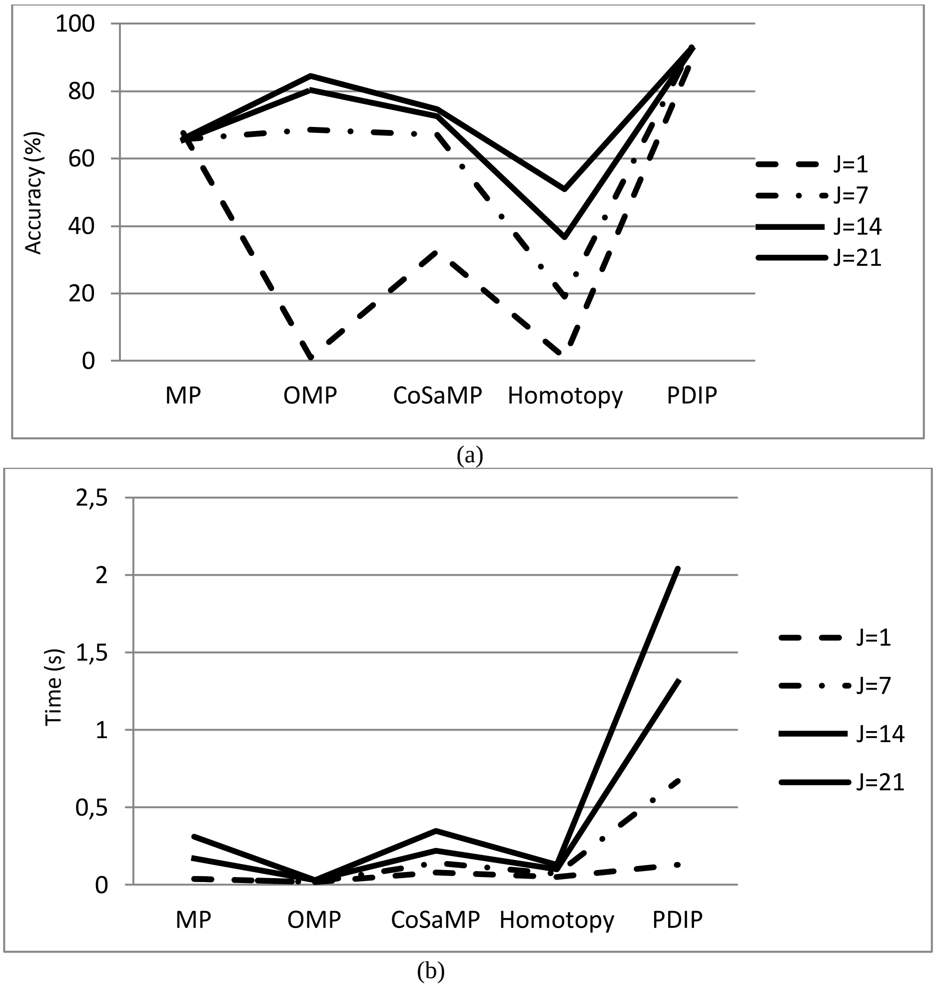


Figure 5. Comparison between (a) accuracy and number of iterations, (b) run time and number of iterations, for five different CS recovery algorithms when different datasets are used for SRC.

REFERENCES

[1] Vaidyanathan, P. , 2001. Generalizations of sampling theorem; seven decades after nyquist. *Circuits and Systems I: Fundamental Theory and Applications*, Vol. 48, No. 9, pp. 1094-1109.

[2] Donoho, D. , 2006. Compressed sensing. *IEEE Transaction on Information Theory*, Vol. 52, No. 4, pp. 1289-1306.

[3] Vo, N. ,Vo, D. ,Challa, S. , 2009. Compressed sensing for face recognition. *Computational Intelligence for Image Processing (CIIP), IEEE Symposium on*, pp. 104-109.

[4] Yang, A. , Ganesh, A. , Ma, A. , 2010. Fast `1-minimization algorithms and application robust face recognition: A review. *Image Processing (ICIP), 17th IEEE International Conference on Image Processing(ICIP)*, pp. 1849-1852.

[5] Candès, E. , Wakin, M. , 2008. An introductuon to compressive sampling. *IEEE Signal Processing Magazine*, Vol. 25, No. 2, pp. 21-30.

[6] Du, L. , Wang, R. , Wan, W. , Yu, X. , 2012. Analysis on greedy reconstruction algorithms based on compressed sensing. *Audio, Language and Image Processing (ICALIP), 2012 International Conference on Image Processing*, Vol. 2, pp. 783-789.

[7] Mallat, S. , Zhang, Z. , 1993. Matching pursuit in a time frequency dictionary. *IEEE Transactions Signal Processing*, Vol. 41, No. 12, pp. 3397-3415.

[8] Asif, M. , Romberg, J. , 2014. Sparse recovery of streaming signals using l1-Homotopy. *IEEE Transactions on Signal Processing*, Vol. 62, No. 16, pp. 4209-4223,

[9] Boyed, S. and Vandenberghe, L. , 2004. *Convex Optimization*, Cambridge Univ. Press, Cambridge, UK.

[10] Candes, E. and Romberg, J. , "l1-magic: Recovery of sparse signals via convex programing", [Online]. Available: http://users.ece.gatech.edu/~justin/l1magic/.

[11] Martinez, M. , "The Ohio State University," 24 Jun 1998. [Online]. Available: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html.

[12] Kuang-Chih Lee, Jeffrey Ho, and David Kriegman, "The extended Yale face database B," [Online]. Available: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.

[13] Spacek, L. , "University of Essex," 20 Jun 2008 . [Online]. Available: http://cswww.essex.ac.uk/mv/allfaces/.

[14] JWright, J. , Yang, A. , Ganesh, A. , Sastry, S. , 2009. Robust face recognition via sparse representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210-227.

[15] Benavente, A. , 1998. The AR face database. *CVC Technical Report 24*.

[16] Lindeberg, T. , 1998. Feature detection with automatic scale selection. *International journal of computer vision*, Vol. 30, No. 2, pp. 79-116.

IMAGE BLOCK COMPRESSED SENSING UNDER LOW SAMPLING-RATIO

Zhengguang Xie, Huang Hongwei and Cai Xu
School of Electronics and Information, Nantong University, Nantong City, China

ABSTRACT

Block Compressed Sensing (BCS) is a new image sampling/compressing method with compressed sensing (CS). To solve the performance degradation of BCS-SPL (BCS with Smoothed Projected Landweber algorithm) at low sampling-ratio, we propose a novel algorithm called Total Variation based Adaptive-Sampling BCS with OMP (TVAS-BCS-OMP). TVAS-BCS-OMP blocks the whole image in an overlapping way to eliminate blocking effect. It assigns sampling-ratio depending on each block' texture complexity, which is measured by the block's Total Variation (TV) so that the blocks with big TV can attain higher sampling-ratio. Then only limited nonzero coefficients in each block are retained according to the adaptively assigned sampling-ratio. At last, we sample the blocks and conducts OMP reconstruction respectively. The experimental results show that under the condition of low initial sampling-ratio (lower than 0.2), TVAS-BCS-OMP achieves better reconstruction precision than BCS-SPL, especially in the blocks with complex texture. In addition, the new algorithm costs shorter reconstruction time than BCS-SPL algorithm.

KEYWORDS

Image reconstruction; Block compressed sensing; Total variation; Overlapped sampling; Adaptive sampling-ratio assignation

1. INTRODUCTION

Compressed sensing(Candes 2006, Donoho 2006, SHI 2009) is a new signal sampling method declared by Candes, Donoho and et al in 2004. It projects the initial signal to a subspace with lower dimension decided by measurement matrix. When the original signal is sparse or compressive and measurement matrix acquires Restrict Isometry Condition (RIC) (Candes 2005), it can be accurately reconstructed through convex optimization methods with few measurement values. As a result, sampling cost is dramatically reduced compared to the Nyquist sampling law. Therefore, CS has a wide applying prospect and has attracted attention all around the world.

CS reconstruction procedure aims to reconstruct signal from measurement values. When CS is applied for a two-dimension (2D) natural image, such problems as poor reconstruction quality, high time costs and high storage space for measurement matrix need to be solved. Till now, several efficient algorithms have been put forward, for instance, Gan's BCS algorithm (Gan 2007) and Mun's BCS-SPL algorithm (Mun 2009). Both of them adopt the Landweber iteration method to reduce computation complexity of image reconstruction. Meanwhile, Wiener filter is used to relieve blocking artifacts. However, unlike Gan's BCS algorithm, BCS-SPL adapts bivariate shrinkage instead of hard shrinkage to provide the requisite sparsiy constraint in the multi-scale decomposition structure. Moreover, several transforms such as contour-lets and complex-valued dual-tree wavelets are considered for their highly directional representation quality in BCS-SPL. Although BCS-SPL can reconstruct an image's basic content under quite few measurements within a quite short time. There exists obvious defects that the Wiener filter and iterative projected Landweber discard some detailed information in the image. As a result, the reconstructed image is quite vague, especially when the sampling-ratio is low. Afterwards, the Self Adaptive Sampling-ratio SPL (SASR-SPL)(ZHANG 2012) extracts each block-texture and employs an adaptive sampling-ratio assigning methodology. Specifically, SASR-SPL estimates each block's sparsity based on the extracting block-texture after basic sampling. Each adaptive sampling-ratio is assigned according to the estimated sparsity so that measuring resources are transferred from smooth blocks to texture ones. Then BCS-SPL is applied to each block under the corresponding assigned sampling-ratio. The reconstruction quality in the texture block is notably improved in this way. Meanwhile, smooth block's reconstruction quality is hardly impaired. Similarly, the algorithm called Sampling Adaptive BCS-SPL based on Edge Detection (SA-BCS-SPL-ED) in (ZHENG 2013.) utilizes gradient to estimate each block's texture complexity and assigns sampling-ratio. Both of SASR-SPL and SA-BCS-SPL-ED can achieve higher reconstruction accuracy compared to BCS-SPL. However, because of assigning sampling-ratio differently for each block and reconstructing one by one in SPL, the time spent for sampling and reconstructing was increased inevitably in BCS-SPL-based algorithms. In order to improve BCS-SPL's performance under low sampling-ratio and efficiently reduce the reconstruction complexity, another adaptive BCS algorithm is proposed in this paper called TVAS-BCS-OMP. The experimental result shows that when the sampling-ratio is below 0.2, the reconstruction quality and computation complexity of TVAS-BCS-OMP are superior to those of BCS-SPL.

2. BLOCK COMPRESSED SENSING

In (Gan 2007), Block Compressed Sensing (BCS) is presented to handle 2D image CS processing. Suppose the original image ($W \times H$) has WH pixels and Q samples. Segment the image into several $B \times B$ blocks. Suppose that x_i is a vector representing block b_i of input image x in raster-scan fashion. The corresponding measured vector y_i is equal to $\Phi_B x_i$. Φ_B is an $m \times B^2$ measurement matrix with $m = \lfloor QB^2 / (WH) \rfloor$, where m denotes the number of uniformly assigned samples in each block. Then the whole measurement matrix Φ for the whole image x is denoted as:

$$\Phi = \begin{bmatrix} \Phi_B & & & \\ & \Phi_B & & \\ & & \ddots & \\ & & & \Phi_B \end{bmatrix} \tag{1}$$

As BCS can independently sample and reconstruct each image block, the calculation complexity is sharply decreased. Then the measurement operator Φ_B is conveniently stored because of its compact size. Last, the encoder doesn't need to send the measured vector of the entire image. Instead, it may send each block once measured, which insures instantaneity.

3. BCS WITH SMOOTHED PROJECTED LANDWEBER ALGORITHM

S.Mun and J.E.Fowler incorporate Projected Landweber algorithm(Piana 1997) into BCS to present BCS-SPL, which proves to possess superior reconstruction accuracy and lower computation complexity. Its main steps are as follows:

Step 1: Segment the image into several $B \times B$ blocks. Suppose that x_i is a vector representing block i of input image x in raster-scan fashion. The corresponding measured vector y_i is $\Phi_B x_i$. Initialize reconstructed image as $x^{(0)} = \Phi^{-1} y$ (set $l = 0$).

Step 2: Smooth the reconstruct image of the l -th iteration as: $\mathbf{x}^{(l)} = Wiener(\mathbf{x}^{(l)})$.

Step 3: For each image block $\mathbf{x}_i^{(l)}$, execute the following iteration:

$$\hat{\mathbf{x}}_i^{(l)} = \mathbf{x}_i^{(l)} + \mathbf{\Phi}_B^T(\mathbf{y}_i^{(l)} - \mathbf{\Phi}_B \mathbf{x}_i^{(l)}) \tag{2}$$

Step 4: Transform the reconstructed image $\hat{\mathbf{x}}^{(l)}$ to Ψ domain. Then get $\hat{\boldsymbol{\Theta}}^{(l)} = \mathrm{Y}\hat{\mathbf{x}}^{(l)}$.
Step 5: Execute thresholding as

$$\boldsymbol{\Theta}^{(l)} = \mathrm{Threshold}(\hat{\boldsymbol{\Theta}}^{(l)}, l) \tag{3}$$

Step 6: Transform $\boldsymbol{\Theta}^{(l)}$ back to time domain $\tilde{X}^{(l)} = \Psi^T \Theta^{(1)}$.
Step 7: For each image block $\tilde{X}_i^{(l)}$, execute the following iteration

$$\tilde{X}_i^{(l+1)} = \tilde{X}_i^{(l)} + \Phi_B^T (y_i^{(l)} - \Phi_B \tilde{X}_i^{(l)}) \tag{4}$$

Step 8: If the termination condition is unsatisfied, then go to Step2. Or else, iteration terminates.
BCS-SPL conducts reconstruction experiments in the domain of several directional transforms such as DCT, DWT and DDWT. The reconstruction results are quite accurate except that the detailed information is somewhat lost due to over smoothing operations. The lower the sampling-ratio is, the vaguer the reconstruction image is poor, especially when sampling-ratio is below 0.2.

4. TOTAL VARIATION BASED ADAPTIVE-SAMPLING BCS WITH OMP

To eliminate the above-mentioned shortcoming of BCS-SPL, we propose the TVAS-BCS-OMP. It mainly includes the adaptive sampling strategy based on TV, BCS overlapped sampling and adaptively setting block sparsity. The framework of TVAS-BCS-OMP is as follows in Figure1.

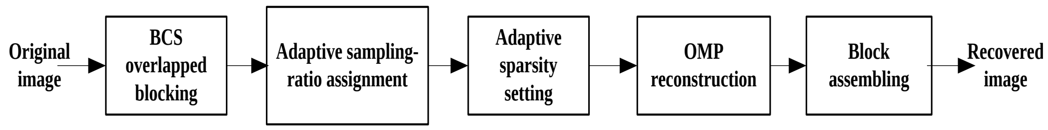


Figure 1. The framework of the algorithm

4.1 TV-based Adaptive Sampling

Essentially, adaptive BCS sampling-ratio is incorporated to assign more sampling resources to texture blocks. The new algorithm uses Total Variation (TV) to estimate the texture complexity of each block. Then the sampling-ratio of each block is decided according to its TV. The specific process is as follows:

Suppose the initial image is of size M×N. Set the initial sampling-ratio as $R \quad (0 < R \leq 1)$. Block the image into $n_B = MN / B^2$ blocks with size B^2. Use $b_{st}(1 \leq s \leq M/B, 1 \leq t \leq N/B)$ to represent the block about to handle. Using P_{ij} to denote pixel locations in the image. TV of block K_{st} is defined as follows in (5).

$$TV_{b_{st}} = \sum_{i=(s-1)B+1}^{sB} \sum_{j=(t-1)B+1}^{tB} \sqrt{\left\| D_{p_{ij}h} \right\|^2 + \left\| D_{p_{ij}v} \right\|^2} \tag{5}$$

ISBN: 978-989-8533-38-8 © 2015

Where $D_{P_{ij}h}$ and $D_{P_{ij}v}$ denote gradient value of P_{ij} in the horizontal and vertical direction respectively(LI 2012). Additionally, similar to, the edge pixels in the neighboring blocks are used to calculate TV of edge pixels in the current block. The sketch for TV calculation is given in Fig.2.

（1）For block b where $1 \le s < M/B, 1 \le t < N/B$,

$$D_{P_{ij}h} = P_{ij} - P_{i,j+1}, D_{P_{ij}v} = P_{ij} - P_{i+1,j} \quad (6)$$

（2）For block b where $s = M/B$ and $1 \le t < N/B$,

$$D_{p_{ij}h} = p_{ij} - p_{i,j+1}, D_{p_{ij}V} = \begin{cases} p_{ij} - p_{i+1,j}, i < sB \\ 0, i = sB \end{cases} \quad (7)$$

（3）For block b where $1 \le s < M/B$ and $t = N/B$,

$$D_{p_{ij}h} = \begin{cases} p_{ij} - p_{i,j+1}, j < tB \\ 0, j = tB \end{cases}, D_{p_{ij}V} = p_{ij} - p_{i+1,j} \quad (8)$$

（4）For block b where $s = M/B$ and $t = N/B$,

$$D_{p_{ij}h} = \begin{cases} p_{ij} - p_{i,j+1}, j < tB \\ 0, j = tB \end{cases}, D_{p_{ij}V} = \begin{cases} p_{ij} - p_{i+1,j}, i < sB \\ 0, i = sB \end{cases} \quad (9)$$

$\overline{TV}$ is the average TV among all blocks, $\overline{TV} = \frac{1}{n_B}\sum_{i=1}^{n_B} TV_i$. Sampling-ratio ri assigned to i-th block is determined by its TV_i, according to Eq.(10).

$$r_i = x[R\mu + R\frac{TV_i}{\overline{TV}}(1-\mu)] \quad (10)$$

Where μ is a controlling factor, ranging within [0,1], to avoid deteriorated reconstruction quality caused by too great difference of sampling-ratio among blocks. Meanwhile, value of ri should be limited to avoid sub-sampling and over-sampling as follows.

$$r_i = \begin{cases} 1; r_i \ge 1 \\ R/r; r_i \le R/r \\ r_i; others \end{cases} \quad (11)$$

γ is a constant which is restricted between 1.1 and 2.4 to determine the minimum sampling-ratio R/g. Then update the value of shrinkage factor x according to (12) in order to guarantee the average of adaptive sampling-ratio equals the initial ratio R.

$$x = \frac{R}{\overline{r_i}} \quad (12)$$

Executing Eqs. (10)~(12) until the value of x satisfies $|x - 1| < 0.001$. Then the corresponding r_i is the adaptive sampling-ratio assigned to block b_i .

4.2 Sampling and Reconstructing of Pixels in Overlapped Part

In order to eliminate blocking artifacts of BCS reconstruction and retain as more detailed information as possible, we employ overlapped sampling and reconstruction. As is shown in Fig.3, the image is blocked in an overlapped way with 2 pixels overlapped. Suppose a block divided by dotted lines has side length of B, then the overlapped block has side length B_overlap of B+1. Fig. 3 shows the value of the additional pixel in the b_1, b_2, b_3 is equal to that of its neighbor pixel.

When reconstructing, the value of pixel in overlapped part is a weighted average of the adjacent blocks. For instance, b_1, b_2, b_3, b_4 get the adaptive sampling-ratio r1， r2， r3， r4 respectively. Their weights are determined by r1， r2， r3， r4, so as to improve reconstruction accuracy of edge pixels. Suppose S1， S2， S3， S4 are respectively the reconstructed values of four pixels in the middle overlapped square of K1， K2 ， K3 and K4 , then their weighted average are determined by Eq. (13).

$$S\_overlap = (r_1 S_1 + r_2 S_2 + r_3 S_3 + r_4 S_4) / (r_1 + r_2 + r_3 + r_4) \tag{13}$$

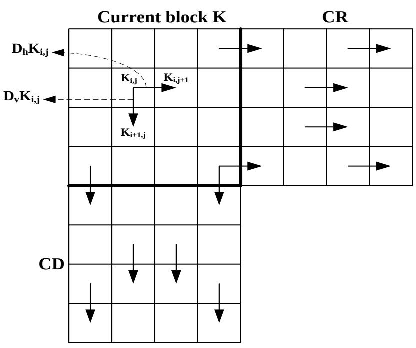


Figure 2. The sketch for TV calculation

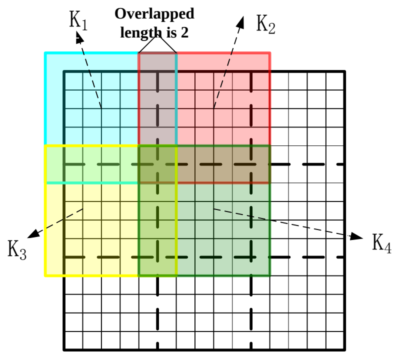


Figure 3. The sketch for overlapped sampling

4.3 Adaptive Sparsity Setting

To ensure that an image is successfully reconstructed in greedy reconstruction algorithm, we employ phase transitions, which are crucial for empirical evaluation of CS recovery algorithms over a wide range of the sparsity level and the measurement number. (Karahanoglu 2013) gives the empirical phase transition curves of OMP when the signal complies with Gaussian, uniform or constant-amplitude random-sign distribution (CARS). It is also pointed out that CARS signal is most difficult for OMP, specifically, CARS signal needs the most measurement number compared to other signal with same sparsity. In the new algorithm, we estimate the functional relationship between sampling-ratio ri and the highest sparsity Ki allowed according to OMP’s phase transition with CARS signal, in order to guarantee 100% successful reconstruction. It is clear that relative sparsity ρi (ρi=Ki/Mi) and ri satisfy linear relation to a large extent, namely, ρi=ari+b. When a, b are set as a=0.4, b=0.1, the discrete points in the phase transition are well matched. Accordingly, the relation between is set as Eq. (14), where $r_i = M_i / N_i$. Afterwards, for each block, only the Ki coefficients with the biggest amplitude are retained, which can not only guarantee successful OMP reconstruction, but also reduce time cost.

$$K_i = (0.4r_i + 0.1)M_i \tag{14}$$

5. EXPERIMENTAL SIMULATION

'Lena' image (512×512) reconstruction results using TVAS-BCS-OMP and BCS-SPL are compared in this section. As done in [5], we employ blocks of size B=32. For TVAS-BCS-OMP, μ in Eq. (11) are set as 0.35 and γ in Eq. (12) equals 1.9. Perform adaptive sampling-ratio assigning after overlapped blocking. Then each block is DCT transformed and retain certain coefficients according to Eq. (14). Number of maximum iterations for OMP is restricted to Mi/2, where Mi is the measurement number assigned to block b_i. The termination threshold is $e = 10^{-6}$. Similarly, BCS-SPL utilizes DCT as sparse transformation and its related parameter are the same as what in [6]. Table 1 lists reconstruct PSNR and time of 'lena' using the two algorithms under different sampling-ratio, where $PSNR = 10\lg\frac{255^2}{MSE}$ (in dB) and time is in s. Recovery simulations are repeated 10 times over i.i.d Gaussian matrix with size $(B+2)\times(B+2)$. For one simulation, measurement matrix for block b_i selects first Mi rows of corresponding measurement matrix. The last result is calculated as the average of the 10 simulation results. In addition, initial sampling-ratio R in Table 1 refers to average sampling-ratio for overlapped blocks of size $(B+2)\times(B+2)$ in TVAS-BCS-OMP. In order to make whole number of measurements in TVAS-BCS-OMP and the non-overlap BCS_SPL_DCT equal, the corresponding sampling-ratio for BCS_SPL_DCT is fixed as $R(B+2)^2/B^2$.

Table 1. PSNR/time of the reconstruction result of image 'lena'

Algorithm	Initial sampling-ratio R						
	0.1	0.125	0.15	0.175	0.2	0.225	0.25
BCS_SPL_DCT	28.17/9.85	29.03/10.47	29.74/8.84	30.44/8.48	31.06/7.48	31.63/7.45	32.16/7.51
TVAS-BCS-OMP	28.67/2.23	29.82/3.27	30.86/4.37	31.80/6.78	32.65/10.52	33.45/16.28	34.18/24.80

Table 2. PSNR/time of reconstruction results of images in Figure 6

algorithm（image）	Initial sampling-ratio R						
	0.1	0.125	0.15	0.175	0.2	0.225	0.25
BCS-SPL-DCT（Fig. 6（a））	21.81/11.17	22.47/9.13	22.88/7.89	23.26/7.00	23.62/7.17	23.96/6.13	24.29/6.58
TVAS-BCS-OMP（Fig. 6（a））	22.63/2.09	23.22/2.93	23.74/3.54	24.24/5.12	24.69/7.09	25.14/10.23	25.56/14.74
BCS-SPL-DCT（Fig. 6（b））	26.13/10.40	26.35/11.31	26.92/10.85	27.41/12.30	28.27/11.42	29.22/10.74	29.24/12.02
TVAS-BCS-OMP（Fig.6（b））	26.95/2.42	28.29/3.39	29.40/5.22	30.44/8.09	31.42/13.23	32.31/21.40	33.16/33.37
BCS-SPL-DCT（Fig. 6（c））	21.68/8.15	21.85/9.16	22.22/8.54	22.45/10.16	22.52/11.44	22.78/10.91	23.34/9.32
TVAS-BCS-OMP（Fig. 6（c））	22.61/2.34	23.10/2.85	23.56/4.14	24.01/5.78	24.45/8.25	24.89/12.48	25.33/18.07
BCS-SPL-DCT（Fig. 6（d））	22.94/10.26	23.31/10.71	23.88/9.43	24.21/10.36	24.74/9.75	25.25/10.01	25.67/10.48
TVAS-BCS-OMP（Fig.6（d））	24.65/2.33	25.76/3.63	26.71/5.37	27.58/9.09	28.36/14.48	29.13/23.54	29.83/36.93

As is clear in Table 1, As R increases gradually from 0.1 to 0.25, TVAS-BCS-OMP provides PSNR value higher than what got by BCS_SPL_DCT all the time, and time cost of BCS_SPL_DCT is roughly constant while that of TVAS-BCS-OMP increases gradually. When R=0.175-0.2, time costs of the two algorithm are roughly equal. Therefore, it is noted that when sampling-ratio R is quite low (R < 0.2), TVAS-BCS-OMP can attain better reconstruction accuracy than BCS_SPL_DCT, meanwhile, time cost of TVAS-BCS-OMP is lower or no higher than the latter algorithm.

The reconstruction results using TVAS-BCS-OMP and BCS_SPL_DCT when R=0.1 and R=0.2 are given respectively in Fig. 4-5. From them, we can conclude that the two algorithms give similar performance towards the smooth areas in the image, but TVAS-BCS-OMP shows better recovery ability in the texture parts. Specifically, when R=0.1, the braid in the recovery image contains many noisy points for BCS_SPL_DCT, while TVAS-BCS-OMP reconstruction result in the same area is relatively smooth. When R=0.2, as can be seen in Fig. 5, the braid and hat stripe in (b) is clearer and more accurate than that in (a), which indicates better reconstruction quality for TVAS-BCS-OMP similarly.

Furthermore, BCS_SPL_DCT and TVAS-BCS-OMP are used for images with richer texture content, which are listed in Fig. 6. The parameter settings are the same as those in ‘lena’ reconstruction simulation and the results are presented in Table 2. From Table 2, it is emphasized that TVAS-BCS-OMP performs better both in reconstruction accuracy and time cost. Time complexity of TVAS-BCS-OMP is O(KMN), so the corresponding time cost is low when sampling-ratio is low. As for BCS_SPL_DCT, the iteration termination condition is quite complex and the time complexity doesn’t have obvious monotonic relations. As M increases, time cost is comparatively smooth and steady, keeping at a quite high level. As a result, TVAS-BCS-OMP has obvious superiority in speed. In a word, TVAS-BCS-OMP possesses higher reconstruction efficiency than BCS-SPL-DCT for texture images.

6. CONCLUSION

In this paper, TVAS-BCS-OMP is proposed based on BCS theory. It can adaptive assign sampling-ratios according to the TV of each image block. Then we estimate the relationship between sampling-ratio and its supported maximum sparsity of the image according to the phase transition of OMP, based on which fixed numbers of nonzero coefficients in DCT domain are retained. CS sampling, OMP reconstruction and restore in an overlapped way to eliminate blocking artifacts are executed afterwards. Experimental results indicate that when sampling-ratio is in a low condition, the new algorithm has better recovery accuracy than BCS-SPL-DCT as TVAS-BCS-OMP does better in texture areas of an image. Meanwhile, TVAS-BCS-OMP costs less recovery time than BCS-SPL-DCT. So TVAS-BCS-OMP better embodies the advantage of CS that accurately reconstruct signal under low sampling-ratio. How to set the initial sampling-ratio according to the restriction of channel bandwidth and apply our algorithm will be the research emphasis later.

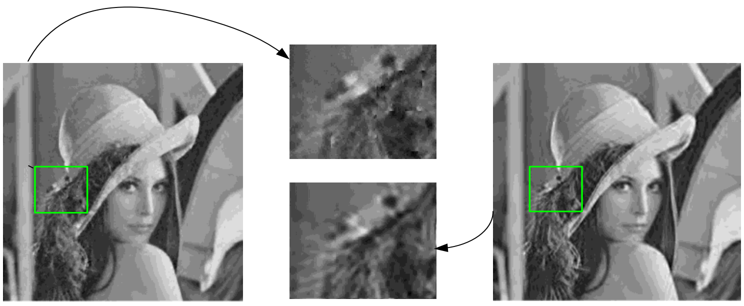


Figure 4. Reconstruction results of BCS_SPL_DCT (left) and TVAS-BCS-OMP (right) when the sampling-ratio R=0.2

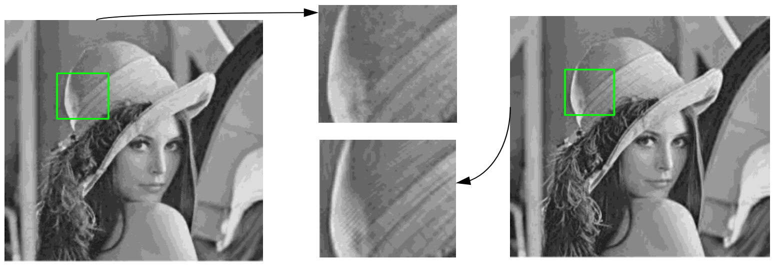


Figure 5. Reconstruction results of BCS_SPL_DCT (left) and TVAS-BCS-OMP (right) when the sampling-ratio R=0.2

(a)bridge.tiff (b)plane.tiff (c)mandrill.tiff (d)Barbara.tiff

Figure 6. The original complex texture images (size: 512×512)

ACKNOWLEDGEMENT

This work was supported in part by the National Natural Science Foundation of China (61171077), the basic research programs of Chinese Department of Transportation (2011-319-813-510), and innovation training program foundation of Nantong University (2009).

REFERENCES

Candes, E. J., Justin R., et al., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on,* Vol. 52, No. 2, pp 489-509.

Donoho, D. L., 2006. Compressed sensing. *Information Theory, IEEE Transactions on,* Vol. 52, No. 4, pp 1289-1306.

Guang-ming, S., L. Dan-hua, et al., 2009. Compressed sensing theory and its research progress. *Chinese Journal of Electronics*, Vol. 37, No. 5, pp 1070-1081.

Candes E J, Tao T., 2005. Decoding by linear programming. *Information Theory, IEEE Transactions on,* Vol. 51, No. 12, pp 4203-4215.

Gan, L., 2007. Block compressed sensing of natural images. *Proceedings of the 2007 15th International Conference on Digital Signal Processing,* pp 403-406.

Mun S, Fowler J E., 2009. Block compressed sensing of images using directional transforms. *Image Processing (ICIP), 2009 16th IEEE International Conference on,* pp 3021-3024.

ZHANG Yang, 2012. Adaptive compressed sensing for natural images. *Xi'an Electronic and Engineering University*.

ZHENG H, ZHU X, 2013. Sampling adaptive block compressed sensing reconstruction algorithm for images based on edge detection. *The Journal of China Universities of Posts and Telecommunications*, Vol. 20, No. 3, pp 97-103.

Piana M, Bertero M, 1997. Projected Landweber method and preconditioning. Inverse Problems, Vol.13, No. 2, pp 441.

LI Kai, ZHANG Shu-fang, LV Wei, et al., 2012. Research on an image block reconstruction algorithm based on TV norm. *Computer Engineering and Applications*, Vol. 48, No. 26, pp 192-196.

Karahanoglu N B, Erdogan H., 2013. Compressed sensing signal recovery via forward–backward pursuit. *Digital Signal Processing,* Vol. 23, No. 5, pp 1539-1548.

Short Papers

CHANGING THE STRUCTURE OF A MOODLE COURSE - A CASE STUDY

Tom Drange[1] and Joachim Kargaard[2]
[1]*Noroff University College*
[2]*Noroff Vocational School*
Elvegata 2A, N-4608 Kristiansand S

ABSTRACT

We are using Moodle as a learning platform in both Bachelor programs and in vocational education offerings, and how learning content is presented through this platform has always been a matter of taste. There are several things to care for in implementing a functional structure of learning content, both from the student's perspective, from the educator's perspective and from the administrator's perspective. We have recently changed the structure of our Bachelor level courses, and with this paper we would like to share our experience with examples and comments from all stakeholders, to provide knowledge in the pursuit of best practices when it comes to the interface that students, educators and administrators are compelled to work with on a daily basis.

KEYWORDS

Online learning, Moodle, content, structure.

1. INTRODUCTION

Todays demand for lifelong learning, driven by the evolution of the Web - followed by content generated by users has created new tools in collaboration and communication (Gomes et al., 2013). Many of these tools are then gathered in Learning Platforms such as Fronter (Fronter, 2015), ItsLearning (ItsLearning, 2015) and Moodle (Moodle, 2015). Moodle is a GPL open-source Learning Platform that we have been using as our learning platform since 2006, but along with the ever-changing world of technology, the features and functionalities of Moodle are increasing and changing at an equally rapid pace. The teaching in itself has been shifting from a traditional "teacher-centred approach" to a more student-centred approach, where the students strive to build knowledge or understanding rather than just gathering facts, which will provide a better experience of the learning process (Shank and Sitze, 2004). Goodwin et al. (2012) suggested that there are three main parts that play an equally important role in e-learning: The Learning Platform, The Course Content and The Interaction. In this paper we will focus on the structure of a course within the learning platform, not its features and elements, in an attempt to structure the learning platform in a way that improves the learning environment for our students, the managing environment for our administrators and the teaching environment for the educators.

2. EXISTING STRUCTURE

Element blocks on each side of the main work area (the middle area) can be somewhat managed by students, so the main focus will be on the structure of the main work area. First of all, the existing structure is per course, which means that each course in all our Bachelor programs has its own course in Moodle. This means twenty-four courses in the Digital Forensics degree program. Each course incorporates the same structure, and it is based on topics with an information area on the top. The information area includes some general resources for the course, such as a general news forum, a Q&A chat, a discussion forum, assignments and deadlines, upload links for the student response to course work and the recordings of all lectures. The number of topics corresponds to the number of modules in each course, ranging from one module to eight modules. Each topic is divided into three sections:

- Lectures: All presentations of lectures in PDF format.
- Additional resources: Additional reading, mini-notes, web resources and so on.
- Activities: Weekly activities and upload links for the student response on weekly activities.

3. CHALLENGES WITH THE EXISTING STRUCTURE

After running this structure for the duration of the first two years we have discovered a few challenges with this structure, and the point of changing the structure is of course to address these challenges:

- **Navigation for students:** When information and content are added to the topics, they become so long that the students must scroll a lot in order to find the desired information. Also, when they progress through their degree program, more and more courses are made available to them, causing the list of available courses to reach a considerable length (twenty-four in total).
- **Students having difficulties finding their grades:** Each course is assessed by coursework and a reflective journal, some of the courses have one coursework, and some of them have two or three coursework. In any case, the grades are contained within each course, so for a student to find one grade, it would be necessary to first find the correct course and then find the correct upload link. There is no method of finding an overview of all grades, since all grades would be course specific.
- **Staff having difficulties figuring out where to publish content:** Different lecturers with different preferences and knowledge of Moodle, publish content in different places, making it difficult for students to find logic in the material published, and for other staff members to figure out what kind of content has been made available to the students.
- **Staff having difficulties finding out where to publish grades:** For the same reasons as the students having difficulties finding their grades in specific courses, the staff also find it challenging, both to find the actual responses and to find out where to publish the grades.
- **Staff having difficulties with getting an overview of a particular student's progress:** Because the marks are listed per course and there is no overview of all the marks, it becomes difficult to track the progress of each individual student. The activity report functions in Moodle is also per course, so in order to track the activity of a particular student in his/her third year of studies, one would have to go through all 24 courses.

4. CHANGING THE STRUCTURE

Several attempts were done along the way to improve on these challenges, such as adding an "assessment course" where all assignments to be graded were implemented and controlled using groups. However, since all teachers had to have access to this course in order to grade and give feedback on assignments, all teachers managed this in their own way. All teachers have their own way of doing things, so the structure of this course spanned slightly out of control, with no logic, making it very hard for both students and staff to get an overview. Since this was a course "on the side" of everything else, teachers often forgot to manage the groups used to provide access for the students. That again led to an increasing dissatisfaction from students not finding their grades. Putting out small fires, as they appear is never a good idea compared to changing the whole structure in order to address all challenges at once, so this is what we did.

The different course formats available in Moodle is Single activity format, Social format, Collapsed Topics, Topics format and Weekly format. We were previously using the Single activity format, creating Moodle activity elements in the order the topics were taught, separated by Labels in order to indicate what topic the activities belonged to. One also has the option of choosing the set structure between Topic, Week, Current Week First, Current Topic First and Day.

The Social format is based on a forum layout, the Weekly format is organized into weekly sections and the only difference between Topics and Collapsed Topics is the possibility to collapse the topics. The Social format does not really fit our purpose - as we need the possibility to create different learning activities to enhance the learning experience for students. The Weekly format is based on the date the course will be started, and since we are starting online courses several times a year - this was not an option. The choice was made to go for the Collapsed Topic, where each course in the program is implemented as a "Topic" in Moodle (Moodle, 2015) to make it easier for students to get an overview without scrolling too much.

We also chose to go for the Set structure of Topic, again because of several starts within the same academic year.

The topics can be collapsed, and the students can have the current topic open to focus on this particular course and thus reduce the need for scrolling. All courses are then divided into four sections:

- A Book element containing practical information about the specific course - such as timetables, reading resources, coursework assignments and deadlines, necessary and important tools that will be used in the course and so on.
- A section with communication and interaction tools such as forums, chats and wiki's.
- A Book element containing all the lectures, activities and resources for the course.
- An assignment section where the students upload responses to the activities and coursework, do quizzes and other deliverable activities.

The new course structure, with the Book element for the course overview, the section with the forum and wiki and the second Book element with the actual learning material make navigation much easier for the students. Now they do not need to scroll to find the desired information. Since each course follows the same structure, everything is kept simple and clear for the students. The structure in the Book element with the learning material also follows the same structure in every course, making it recognizable for the students and therefore provides easy navigation.

The element is divided into modules as chapters in the book. The modules follow the approved study plan for the course, and it is further divided into four sub-chapters:

- Lecture Presentations: The lecturer of the specific course uploads all presentations in a PDF format beforehand, so that students can have a look at them before the actual lecture. This makes the students prepared for the lecture and they have often noted questions they want answered during the lecture. If the lecturer does not provide the answers during the lecture - we use a Skype chat that all students enrolled in the course follow in addition to all staff connected to the course. The students can ask questions through this chat in order to follow up on material that is unclear.
- Recordings: We provide both synchronous and asynchronous online learning, so in addition to streaming all lectures live, we also record them so students can look at them at any time and as many times as they need to in order to grasp the topics presented.
- Activities: The learning activities related to that module are found here.
- Resources: Additional resources such as links to websites, interesting and related talks, tools, mini-notes and anything useful that does not fit into the previous three sub-chapters.

We have also a section named "WELCOME TO NUC - PRACTICAL STUDENT INFORMATION" containing inspiring welcoming words in addition to information that students often requests, such as course plans, teacher/tutor support, forum guidelines and a student handbook with tips and tricks for the course.

5. HAVE WE ADDRESSED THE CHALLENGES BY CHANGING THE STRUCTURE?

How did the change in structure address the issues that had been identified?

- **Navigation for students:** Having the entire BSc course in one Moodle course has made navigation much easier for students. It is now possible to find relevant course material easily and quickly. We also avoid students having to scroll a lot to find content. Using the same structure for the entire course also ensures that it is kept simple and makes it easier for the students to grasp what is going on.

- **Students having difficulties finding their grades:** Because the elements are very clearly defined and uniform across the courses, it is easy for the students to find out where they need to find their grades. We also use the "Welcome to NUC" section to explain to students where they can find all the necessary information, including where they can find their grades.
- **Staff having difficulties finding where to publish content:** All elements now have the same structure with clearly defined sections for content. This makes it easy for staff to publish content, even if they are not familiar with the system beforehand. The logical structure present in all course elements guides both existing and future staff to where different types of content should be published.
- **Staff having difficulties finding where to publish grades:** The same arguments are valid for this challenge, namely that the clearly defined and uniformed way of structuring the elements, makes it easy for staff to both find the actual assessments and publish the grades and feedback.
- **Staff having difficulties with getting an overview of a particular students progress:** Since all information now is gathered in the same course, staff can use several functions to get an overview of both each individual student and of all students.

6. SUMMARY AND CONCLUSIONS

The structure of the BSc course was changed to ensure that students' could navigate the course more easily and through that create a better learning environment. Swan (2001), citing a number of authors, identifies course design as a critical factor in determining quantity, quality, and type of interactivity (learner interaction with content, instructors or peers) in a course (Brindley et al. 2009). Alves et al. (2013) found statistically significant correlations between the variety and quality of Moodle resources and students' results. Effective course design minimizes course navigation time, leaving more time for students to spend collaborating, communicating and engaging with the course (Teaching at UNSW, 2013). Also one needs to consider the lack of experience among most of the educators. This is according to Bagiati (2014) a unique challenge when developing online courses as opposed to developing traditional courses. The main structure and feel of the courses cannot be left for each individual educator to decide, when students expects consistency and quality (Puzziferro & Shelton, 2008). It is also important to bear in mind that one can not expect students to automatically know how to learn online, just as little as one can expect teaching staff to automatically know how to teach online (Palloff & Pratt, 2001) so education for both students and staff on the usage of the online tools, is necessary in order to get the most out of the online course structure. It must however be an ongoing dynamic process with the learning outcomes of the student as the main focus, and we as educators must of course continue to support new methods of learning design in order to prevent online course development from ceasing to move forward (Wright, 2015).

7. FURTHER WORK

The structure currently used in our vocational programs is the same as the one used previously in the Bachelor programs. We are changing this to the new structure in August of this year, which means we will be able to conduct surveys among the students using the old structure towards the summer, and then conduct a survey among the same students after the change. This will give us a pointer as to the level of success we have achieved by changing the structure. A small, unofficial survey on how students and staff in the bachelor programs like the new structure, has indicated that 45% of the students are positive to the change, and 25% of the students do not have an opinion. 100% of the staff is positive to the change. This indicates that there are grounds to continue the work that we have started.

REFERENCES

Alves, G.R., et al, 2013. Impact of Different Moodle Course Designs on Students' Performance, *iJep, v3, Special Issue 2*

Bagiati, A., 2014. Developing an Online Course: Challenges and Enablers, *SEFI 42nd Annual Conference, Birmingham, UK* [Internet], Available from http://www.sefi.be/conference-2014/0030.pdf (Accessed 30 January, 2015)

Brindley, J.E., et al, 2009. Creating Effective Collaborative Learning Groups in an Online Environment, *The International Review of Research in Open and Distributed Learning, v10 n3*

Fronter, Learning Platform [Internet], Available at: http://com.fronter.info/ (Accessed 10th March, 2015)

Gomes, N. F., Fonseca, C., and Serrano, M.J.H., 2013 Building a Successful e-Learning Project in Higher Education, *GSTF Journal on Computing (JoC) Vol.3 No.1, March 2013*

Goodwin, M., et al, 2012. Assessing the accessibility of e-learning, NOKOBIT 2012, *Universitetet i Nordland 19-21. November,* s 145-158.

Itslearning, Learning Platform [Internet], Available at: http://www.itslearning.eu/ (Accessed 10th March, 2015)

Moodle, Open Source Learning Platform [Internet], Available at: http://moodle.com/ (Accessed 30 January, 2015)

Palloff, Rena M., Pratt, K., (2001), "Lessons from the Cyberspace Classroom", *17TH Annual Conference on Distance Teaching and Learning*, Available at: http://www.uwex.edu/disted/conference/resource_library/proceedings/01_20.pdf (Accessed 12 March, 2015)

Puzziferro, M., Shelton, K., 2008. A Model for Developing High-Quality Online Courses: Integrating a Systems Approach with Learning Theory, *Journal of Asynchronous Learning Networks*, v12 n3-4 p119-136

Shank, P., Sitze, A., 2004. *Making Sense of Online Learning: A Guide for Beginners and the Truly Skeptical*, Pfeiffer 978-0787969820

Swan, K., 2001. Virtual Interaction: Design Factors Affecting Students' Satisfaction and Perceived Learning in Asynchronous Online Courses, *Distance Education*, v22 n2 p 306-311

Teaching at UNSW, 2013. [Internet] https://teaching.unsw.edu.au/moodle-course-design (Accessed 30 January, 2015)

Wright, P., 2015. Comparing E-tivities, e-Moderation and the Five Stage Model to the Community of Inquiry Model for Online Learning Design, *The Online Journal of Distance Education and e-Learning, Volume 3, Issue 1* [Internet] Available from http://www.tojdel.net/pdf/v03i01/v03i01-01.pdf (Accessed 30 January, 2015)

ISBN: 978-989-8533-38-8 © 2015

SIMULATION OF LOW-VISION EXPERIENCE BY USING A HEAD-MOUNTED VIRTUAL REALITY SYSTEM

Yohsuke Yoshioka
Chiba University (JAPAN), Department of Architecture, 1-33, Yayoi-cho, Inage-ku, Chiba, Japan

ABSTRACT

In this study, a system for simulating low-vision experience was developed by using immersive virtual reality technology. The proposed system consisted of a wide-view head-mounted display and an eye tracker that could modify the display in concert with real-time fixation patterns to restrict an arbitrary area of the human visual field. We conducted an experiment to evaluate the validity of the proposed system. Subjects were asked to walk through a small maze under two visual conditions in which 10 degrees of their central visual field was restricted artificially with black circles using the proposed system. Under one of the visual conditions, the restriction moved synchronously with the subjects' eye movement so that the 10 degrees of the central fovea was always restricted despite eye movement; on the other hand, under the other condition, the restriction could not move. The experiment results indicated that the time for walking through the entire maze under the visual condition with 10 degrees of the central visual field restricted in synchronization with eye movement was longer than that under the condition in which 10 degrees of the fixed central area of the screen were restricted. Further, the amplitude of the walking trajectory on a straight pass under the visual condition with 10 degrees of the central visual field restricted in synchronization with eye movement was higher than that under the condition in which 10 degrees of the fixed central area of the screen was restricted. The influence of the eye-following movement of the restricting black circles and the validity of the proposed system were investigated.

KEYWORDS

Immersive Virtual Environment, Head-Mounted Display, Eye Movement, Low Vision, Restriction of Visual Field.

1. INTRODUCTION

Low vision is the partial loss of eyesight that makes carrying out daily activities difficult, such as reading, recognizing faces, or walking through a crowded place. Every year, billions of people in the world lose part of their eyesight because of eye diseases or health conditions such as eye injuries, macular degeneration, cataract, diabetes, and glaucoma.

Because the disorder is caused by the loss of functionality of a part of the visual field, regular glasses, contact lenses, or medicine cannot essentially resolve the difficulties for many patients suffering from low vision. These patients are required to find methods to view the world with their other sensory organs, which is very effort-intensive. It is important to maintain their quality of life through vision rehabilitation methods, and similarly, it is important to arrange and design the daily environment of these patients in a suitable manner. Some popular examples of such designs are environments that are sufficiently bright, and signs and tags that are large and clear.

However, the most crucial issue with current low vision care systems is that people with normal vision find it difficult to imagine the impact of the loss of even a small vision field. We develop a system to simulate a low vision experience that uses the leading-edge technology of immersive virtual environment. The proposed system consists of a wide-view head-mounted display (HMD) and an eye tracker for restricting arbitrary areas of the human visual field.

A person with normal vision can use the proposed system to experience low-vision eyesight. Further, the proposed system would enable researchers in the field of low vision care study to examine the influence of partial loss of a specific area of the visual field for many different individuals several times. The system could help create a procedural breakthrough in this field, because it is almost impossible to find low-vision participants that suffer from the loss of the exact same areas of visual fields for research. Specific behaviors of individuals under low-vision conditions could be analyzed to improve the environment for individuals with low vision.

2. SYSTEM

The proposed system consists of a wide-view HMD (Nvis: nVisor SX111), a position-tracking system (WorldViz: PPT Optical Tracker), an inertial orientation reference system (InterSense: InertiaCube2+), and an eye tracker (Arrington Research: Binocular Eye-tracking System).

The precise real-time position and orientation of the subject's head in real space was tracked with the position-tracking system consisting of eight high-resolution cameras and the inertial orientation reference system arranged on the back of the HMD. The virtual space displayed on the HMD was linked with the calculated real-time position and orientation of the subject in real space so that the subject could walk and look around the virtual space. The wide-view HMD has two high-definition display screens, each of 1280 × 1024 pixels, for each eyeball, which allows subjects to view the diagonal angle of 111 degrees of the visual field through binocular vision and stereopsis.

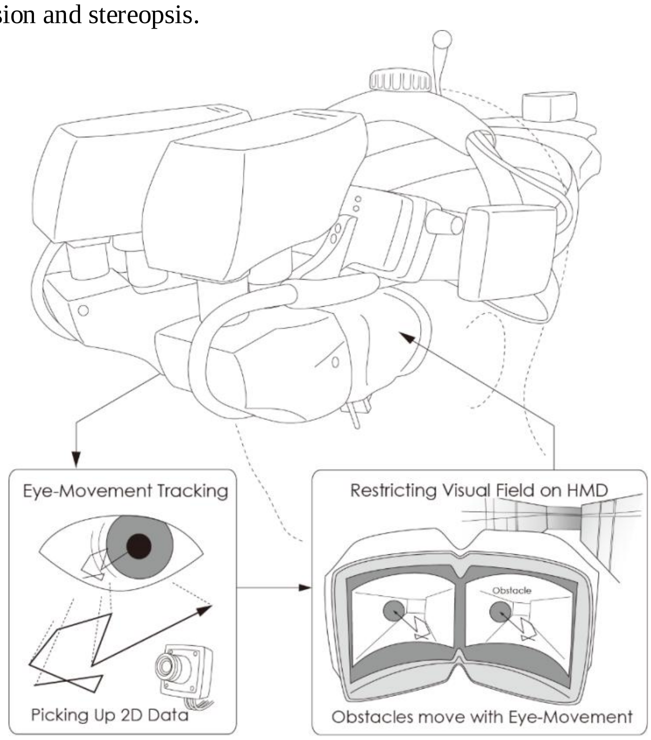


Figure 1. Arrangement of HMD and eye tracker in the simulation system

A feature of the system is that it also has a binocular eye-tracking system in the HMD (Figure 1). Small-sized high-precision cameras and mirrors were combined and installed in the gap between the HMD screen and eyeball in order to detect the eye movement of the subject. The detected eye movement was transmitted directly to the workstation connected with the HMD, and the precise position of each subject's fixation could be calculated through software simultaneously.

Vizard 4.0 (World Viz) software was used to describe the virtual world on the screen of the HMD and for combining the three sequential data transmitted from the three arranged devices, i.e., subject's position, subject's orientation, and subject's eye movement. In particular, by using the real-time fixation position, the software could modify the display in concert with the subject's fixation pattern on the virtual world. For example, we could display small black shields in the virtual world, which were controlled in synchronization with the eye movement in real-time to restrict the view to a specific area of the visual field.

3. EVALUATION

The proposed system allowed normal-vision individuals to experience low vision by displaying obstacles of specific shapes and textures, black circles, and semitransparent small holes that moved in synchronization

ISBN: 978-989-8533-38-8 © 2015

with eyeball movement. To evaluate the validity of the proposed system, we conducted a short experiment with twenty subjects without any visual disorders. The subjects' central visual fields were restricted with small synchronously moving black circles. Written informed consents were obtained from all subjects for publication of this case report and accompanying images.

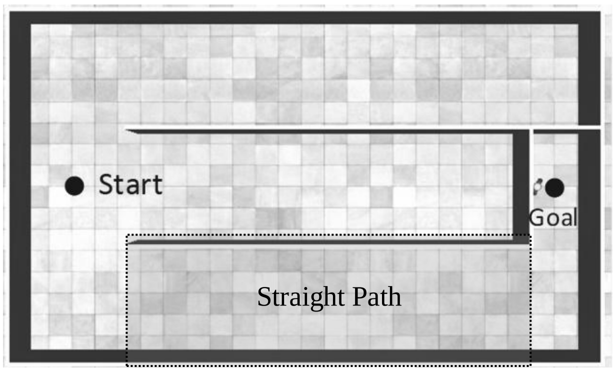


Figure 2. Maze for the evaluation experiment

Figure 2 shows a plan of a short virtual maze that the subjects had to walk through in our evaluation experiment. The maze was arranged within an area of 3 m × 5 m with 3-m high virtual walls; all the corridors in the maze were 1-m wide. We designed a three-way intersection at the starting point of the maze as well two dead-ends in order to examine the behavior of subjects in each situation under restricted visual conditions. The maze also had one floating icon at the goal position, as shown in Figure 2. To complete the trial, the subject was asked to find and touch the icon. The virtual walls had a wood paneling-like texture and the floor of the virtual world was covered with small gray tiles.

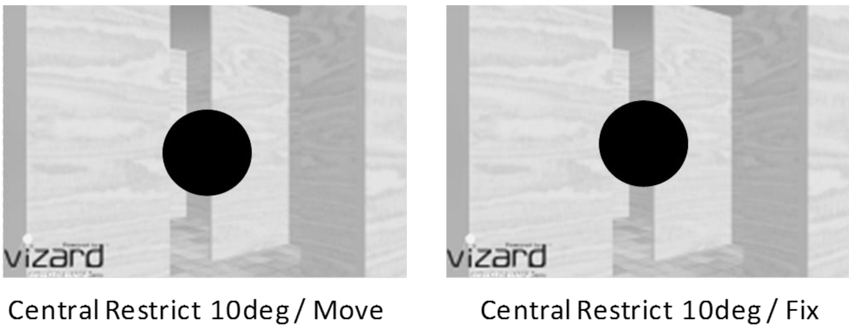


Figure 3. Visual conditions

Subjects were asked to walk through the short virtual maze from the starting point to the goal twice under the following two visual conditions (Figure 3). The effects of the eye-following movement of the circle could be examined by comparing the subjects' walking speed or walking trajectory under the two visual conditions.

- **Central Restrict 10 deg/Move:**
  The subjects' central visual fields were restricted with small black circles that moved according to the eye movement. The size of the circles could cover an area of 10 degrees around the real-time fixation points.
- **Central Restrict 10 deg/Fix:**
  Virtual black circles of the same size as in the "Central Restrict 10 deg/Move" condition were fixed on the center and they did not move with eye movement.

4. RESULTS

Figure 3 shows the average time to walk through the entire maze under the two visual conditions. Figure 4 shows the average amplitude of the walking trajectory in a straight path as indexed in Figure 2 under the two visual conditions.

The P values were assessed to indicate the differences in mean values using a Bonferroni multiple comparison procedure; two significant differences could be found between the two visual conditions.

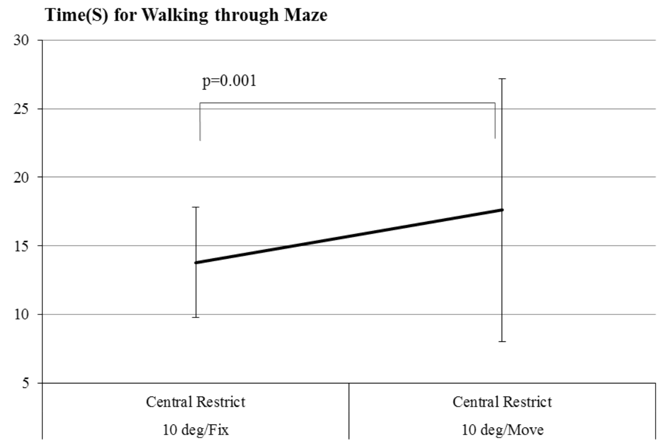


Figure 4. Walking time under the two visual conditions

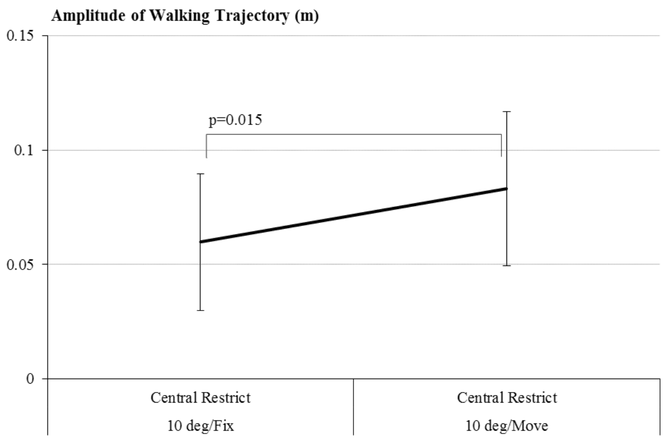


Figure 5. Amplitude of the walking trajectory in a straight path

The first difference is that the time taken to walk through the entire maze under the visual condition with a moving restriction was significantly longer than that under the visual condition with a fixed restriction (p=0.001). The second difference is that the amplitude of the walking trajectory on a straight pass under the visual condition with a moving restriction was significantly higher than that under the visual condition with a fixed restriction (p=0.015).

These results indicate that the walking task under the visual condition with a moving restriction is more difficult than that under the visual condition with a fixed restriction. Further, since it would be difficult to confirm the three-dimensional details of the end of the straight path as the route to the destination under the moving condition, the subject might shift his or her viewing point to the right and left by walking on a meandering line to enable a motion parallax that facilitates a sense of three-dimensional distance.

5. CONCLUSION

In this study, a system that simulates low-vision experience was successfully developed by using virtual reality technology with a wide-view HMD and an eye tracker. The following are the results of the evaluation experiment conducted:

- The time for walking through the entire maze under the visual condition with 10 degrees of the central visual field restricted in synchronization with eye movement were significantly longer than the time under the condition in which 10 degrees of the fixed central area of the screen was restricted.

- The amplitude of the walking trajectory on a straight pass under the visual condition with 10 degrees of the central visual field restricted in synchronization with eye movement was significantly higher than that under the condition in which 10 degrees of the fixed central area of the screen was restricted.

Both of the significant differences resulted from the difference in the two visual conditions, i.e., the central visual restriction moves/does not move with the subject’s eye movement.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI, Grant-in-Aid for Young Scientists (A) No.22686056, Grant-in-Aid for Scientific Research (B) No. 26289210.

REFERENCES

1. Anstis, S. M. 1974. A chart demonstrarin variations in acuity with retinal position. *Vision Research*, Vol.14, No.7, pp.589–592.
2. Blanchard, H. E. et al, 1984. Time course of visual information utilization during fixations in reading. *Journal of experimental psychology*. Vol.10, No.1, pp.75–89.
3. Brown, B., 1972. Resolution thresholds for moving targets at the fovea and in the peripheral retina. *Vision research*, Vol.12, No.2, pp.293–304.
4. Ikeda, M., & Saida, S., 1978. Span of Recognition in Reading. *Vision Research*, Vol.18, No.1, pp.83–88.
5. Johannson, G., 1977. *Visual perception of locomotion elicited and controlled by a bright spot moving in the periphery of the visual field. Report - Depertment of Psychology*, University of Uppsala.
6. Mateeff, S., & Gourevich, A., 1983. Peripheral vision and perceived visual direction. *Biological cybernetics*, Vol.49, No.2, pp.111–118.
7. Osaka, N., 1990. Peripheral lower visual fields: A neglected factor?. *Behavioral and Brain Sciences*, Vol.13, No.3, pp.555. Cambridge University Press.
8. Polyak, S. L., 1941. *The retina: the anatomy and the histology of the retina in man, ape, and monkey, including the consideration of visual functions, the history of physiological optics, and the histological laboratory technique*. University of Chicago Press.
9. Previc, F. H., 1990. Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. *Behavioral and Brain Sciences*, Vol.13, No.3, pp.519–542.
10. Saida, S., & Ikeda, M., 1979. Useful visual field size for pattern perception. *Perception & psychophysics*, Vol.25, No.2, pp.119–125.
11. Ungerleider, L. G., & Mishkin, M., 1982. Two cortical visual systems. *Analysis of Visual Behavior*, Vol.549, pp.549–586.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

GAMIFIED PLATFORM TO SUPPORT CHILDREN WITH OBESITY

Yeray del Cristo Barrios Fleitas, Carina Soledad González González,
Eduardo Aníbal Lalla Ruiz and Pedro Antonio Toledo Delgado
Department of Computer and Systems Engineering, University of La Laguna

ABSTRACT

The ProViTao project is an initiative aimed at helping children with obesity problems to overcome this disease in a fun and effective way through the use of gamification. In this project, during an obesity treatment of nine months, only the first three are supervised by a medical team into the hospital. After that period, the treatment is relegated to the family for the next six months. This may involve some disadvantages for monitoring and supervising the patients. Hence, the goal of this work is to provide a platform, ProViTao APP, for monitoring and supervising the patient in those six last months of treatment, which allows to improve the connection between the medical team and the patient and family. ProViTao consists of a mobile and web application, both similar and with multiple functions. This platform covers all the aspects and processes of the treatment. Namely, (i) store and sync information in the cloud, (ii) a user manager and file system. (iii) tools for analysis and data collection, (iv) a communication channel between users, and (v) sports module.

KEYWORDS

Gamification, human-computer interaction, childhood obesity, e-health.

1. INTRODUCTION

One of the most important risks of childhood obesity is that they tend to remain obese in adulthood while also increasing the likelihood of no communicable diseases (see Whitaker et al., 1997; Must and Strauss, 1999; Freedman et al., 2001). "The program of active video games for the treatment of obesity and diabetes" (ProViTao) is a project aimed at supporting the treatment of childhood obesity at early ages by means gamification (i.e., the use of games for motivating and achieving goals).

The obesity treatment intervention contemplated in this work takes nine months and it is intended to gradually bring the patient over a habit of healthy living. This period is divided into two parts. The first part last three months and along it the child is tutored directly by the medical team. The second part of the intervention takes the remaining six months and the supervision of the medical team is performed through telephone interviews. This may lead to various problems related to their age and the capacity of the patient to change their habits, compromising the treatment. To address these shortcomings, we propose ProViTao APP, an environment that supports the intervention through various training and monitoring tools to improve patients' adherence to the program as well as enhancing their learning in a fun way using gamification techniques. In this regard, the contribution of ProViTao to the existing tools is based on the way it adapts itself to the user, leading to a more rewarding user experience which is translated into an adhesion of healthy lifestyles.

The goals of this work are, on the one hand, to describe the design and implementation of the proposed technological platform, ProViTao APP, as a support tool for the treatment of childhood obesity. On the other hand, to depict the basics of the platform while discussing some strategies and uses of gamification to the obesity treatment.

The remainder of this paper is organized as follows. Section 2 describes the proposed platform, ProViTao. Afterwards, Section 3 briefly reviews some strategies and gamification appliances to tackle childhood obesity. Finally, Section 4 draws the concluding remarks extracted from the work and indicates several promising directions for further research.

ISBN: 978-989-8533-38-8 © 2015

2. PROVITAO APP

One of the main features of ProViTao APP is the automation of processes and intelligent adaptation of interfaces, blended with gamification components.

Figure 1. ProViTao APP Interface

The interface of ProViTao APP is designed to behave differently depending on the user profile. This way, depending some input data related to the user such as the age and the role, the user may have access to different features and information of the application. Thus, for example, a user with the role of a doctor will have access to patient panel and information in contrast to a user with the patient role that instead of it will have access to the gaming platform.

The types of user roles allow to enjoy the same application without interfering among them. In Figure 2, we depict the four possible roles for which the application is designed. Namely, patients (i.e., children with obesity problems following the treatment exposed in this work), family (i.e., parents or legal guardians), medics (i.e., doctors, nurses and nutritionists involved in the project, and staff (i.e., analysts and researchers).

Figure 2. Possible roles within ProViTao APP

Furthermore, ProViTao APP not only manages the visuals to suit your role, but uses the remaining information hosted in the profile to automate as many tasks -related to the treatment- as possible. This way, the information contained in the user profile is used to infer some information when crossing the data from the profile with the one obtained from a test. It should be mentioned that the use of the user data seeks to generate in the user the feeling that the application has been designed exclusively for him / her.

Researchers and analysts involved in the development of the platform use it mainly to monitor the behavior of the patients and obtaining some knowledge from it. That is why ProViTao APP offers several tools for bidirectional communication between users. Based on previous ProViTao interventions the actual model considers two communication tools (see Figure 3) a chat adapted for children and a generator questionnaires. Both tools allow the direct interaction among patients and doctors, bridging the gap that separates them.

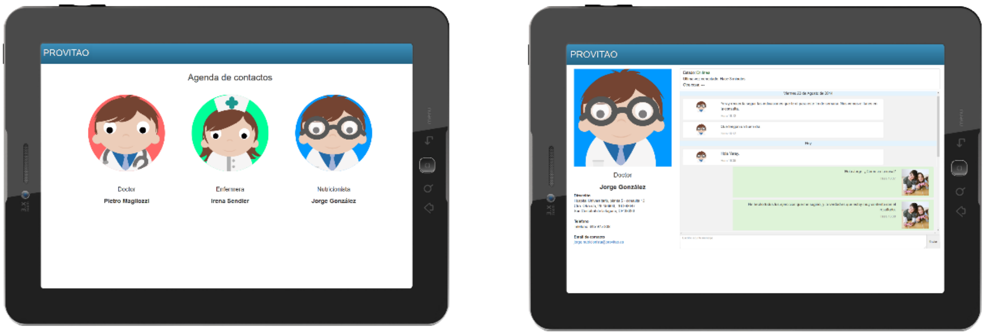


Figure 3. ProViTao communication Tools

3. GAMIFICATION AND VIDEOGAMES

One of the most favorite activities children want to do is to play, and, hence, taking advantage of it may be the most appropriate way to help them acquire healthy habits. This way, ProViTao takes the concept of play and bundle it with the goal of improving the habits of the children allowing them to learn through video games, but also making the environment a pitch.

In this work we propose two training games that have been developed in ProViTao to promote healthy lifestyles.

- The first is called "The Account Steps" and consists of a pedometer with purpose. The doctor assigns a task type exercise jogging or running, with a certain distance to a particular child. This distance is structured in a route divided into several separate control points. Each time control point indicate that a part of the section has been completed, thus, the child should not go back. The route is surrounded by a fantastic narrative where the child is hunter, with the mission to recover all the dragons that escaped from a castle. Each step he takes brings him closer to capture the dragon (i.e., the control point). If the child decide to leave before reaching the dragon, then he/she may lose all the points collected from the last checkpoint to that moment. Once all the dragons are captured, that is, all the control points have been visited then a final dragon can be captured by answering one question provided by the application. If the child is successful in this last part, he/she will receive extra points and a different medal (value, consistency, etc.).

- The second game is called "treasure hunt" and consists of an application that uses geo-location to place the user at the center of a Google map. The player must go in search of attractions related to healthy food, sports or any other healthy item. Each marked location is transmitted to other users who can validate that it is indeed a source of concern. The more people validate the location, the more points generates this achievement on your profile. In the end, it will create a common map of attractions for the acquisition of healthy lifestyles.

ISBN: 978-989-8533-38-8 © 2015

Figure 4. ProViTao videogames for promoting healthy habits

Beside these games, an additional system of ranks and points related to the application motivates the patients to upgrade their level. The points can be increase by means of the time spent using the software, achieving merits or by the interaction with other users (e.g., medics, familiars). Players who accumulate enough points can level-up its category and acquire a new rank. Although the patient is encouraged to strive to win, he has not wanted to do much emphasis on the display of an overall score to avoid discouragement. Instead, a series of positive messages motivate the user to be better every day.

4. CONCLUSION

In this paper, we propose a platform gamified intended to support the treatment of childhood obesity and diabetes. It is based on process automation, intelligent adaptation of interfaces, and game mechanics for achieving an ideal mix that motivate healthy lifestyle habits in children. In this regard, in this work we have proposed ProViTao APP as a support tool for medical teams when they have to virtually follow-up and monitor children. On the other hand, the gamification included in ProViTao APP promotes a treatment adherence and a progress in the weight loss maintenance through the uses of games.

On the basis of the findings presented in this work, the next stage of our research will be focused on an intensive analysis of the benefits provided by ProViTao on patients as well as its integration with further developments in the spirit of video games for training in healthy lifestyles. On the other hand, we will conduct the testing with real users during the 2015/2016 intervention in order to validate our proposal. Moreover, another feature that will be included as future improvement of the application the adaptability of the interface for those users with visual impairment. Following the philosophy of "full automatic", the software will change the layout, content or form of the interfaces to be adapted to the user disability.

REFERENCES

Whitaker, R. C. et al, 1997. Predicting obesity in young adulthood from childhood and parental obesity. New England Journal of Medicine, Vol. 337, No. 13, pp. 69-873.

Must, A., and Strauss, R. S., 1999. Risks and consequences of childhood and adolescent obesity. International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity, Vol. 23, pp 2-11.

Freedman, D. S., Khan, L. K., Dietz, W. H., Srinivasan, S. R., & Berenson, G. S. (2001). Relationship of childhood obesity to coronary heart disease risk factors in adulthood: the Bogalusa Heart Study. Pediatrics, Vol. 108, No 3, pp. 712-718.

PLUG: A ZIGBEE NETWORK FOR PLAYING UBIQUITOUS GAMES

Alexandre Topol, Jérôme Dupire and Viviane Gal
CNAM-CEDRIC, 292, rue Saint Martin, 75003 Paris, France

ABSTRACT

In this paper we present the project PLUG (PLay Ubiquitous Game and play more) the aim of which is to deploy a ubiquitous game in a museum. For this, we required wireless communication between players and the game server, and the deployment of a wireless sensor network for monitoring the physiological activity and movements of the players. Among potential technologies, we selected zigbee. Here we describe the project and the challenges of devising a cheap and useful wireless network solution, including our experiences dealing with hardware and software issues. This work can be used as a reference for future game projects.

KEYWORDS

Ubiquitous, pervasive, video game, wireless sensors network (WSN), physiology, mobility, entertainment.

1. INTRODUCTION

With video games, we can produce feelings of immersion in a variety of ways, but it is interactivity that differentiates video games from passive activities. Voluntary actions of players are already fairly well integrated through human machine interfaces (e.g. keyboard, Wiimote). Non-voluntary actions, however, such as physiological or emotional variations, are just beginning to be studied.

The Plug Project - PLUG (PLay Ubiquitous Game and play more) studies the use of embedded and mobile technologies for building pervasive and ubiquitous games, as well as player acceptance of such games from socio-cultural, economic and industry points of view.

The Cnam museum *(Museum of Arts and Crafts)*[1] was open to innovative ideas for attracting children and teenagers. Our choice of a pervasive game[9], inspired by the game “Happy Families”, was a natural fit, that one enriched with puzzles. In this game, players collect cards pertaining to the four families of museum objects. Using the NFC facility of a mobile phone, players collect virtual cards representing real objects of the museum. The virtual cards (RFID tags) are located near the real objects chosen for the game. The virtual universe of the pervasive game is filled with animated objects with their personalities attributed to them by their inventor. Once they are removed by the player by moving a real object moved toward a phone and vice versa, the player hears a text that provides feelings and emotions related to the inventor or to the object itself. The game is also competitive since the game’s winner is the team player that gets the highest score within a 55 minute game session.

To ensure the project’s success, mobile telephony, WIFI, Bluetooth, RFID were studied. Our main concern was the design of pervasive games using wireless technology. In this context, the use of traditional interactions is difficult due to movement. We used physiological monitoring as a means to interactively adapt the narration to the emotional state of the player using "embodied" sensors that monitored heart rate and activity level.

The project included ten academic and industrial partners. In addition to the completely autonomous WSN described hereafter, the game exploited two other technologies: RFID to collect objects, and mobile touch screens needed for player interaction. For want of a single communication system that could manage both these elements, we used GSM NFC and the iPod touch™ systems connected via Bluetooth. Communication between the game server and the gamers used both these systems.

1 http://www.arts-et-metiers.net/musee/visitor-information

In order to detail the physiological state of the player, we attempted to retrieve, analyse and transmit the physiological player's data to the game server. This was in order to modulate game situations by modifying the narration in real-time. The player's state has an impact on the rules of our game. The game's behaviour was changed by a human game master or an algorithm. In order to collect physiological signals that could play an important part in the detection of emotion, we created and/or used various sensors: accelerometer for the movements, electrocardiogram (EKG) to compute the heart rate, temperature, breath rate, galvanic skin response. In practice, only the accelerometer and the EKG were integrated in the final game design because the temperature signal suffered from inertia, which made it useless in our context. The breath rate signal was too noisy, due to the permanent motion of the player, and the galvanic skin response sensor appeared to be disturbing the game device handling.

Sensor - Sensors are connected to the analogical inputs of the microcontroller, reading values in the form of electric potentials and converting them into numerical values, filtering those values if needed, and concatenating them and constructing a message which is sent via a serial port. The connected radio module transmits the message which is conveyed through a WSN to the central server. The game is managed on the server using an adaptability engine that takes into account the physiological values received. The process allows us to obtain the physiological state of a player within a few milliseconds.

Our work relates to manufacturing and operating the various sensors were described in previous articles[1, 2]. This project focuses on the WSN transmission of physiological signals sensed on a free and moving player.

In order to minimize load, the network transmitting the physiological signals needed to be independent from the main network of the game. The major difficulty came from the institutional context of the game's location. It was formally impossible to install equipment for two reasons: architectural (the buildings are old and protected) and, regulation and safety. The challenge can be summarized in the deployment of a WSN that covers all the rooms involved in the game - around 6000 m^2 - with light and invisible installation.

In this rest of this paper, we review the literature on pervasive games then compare wireless technologies and describe the various zigbee implementations and the experiment in section 3. We conclude with a summary of our contribution.

2. STATE OF THE ART

Ubiquitous computing lead to an increase in work on pervasive games. In Memoriam[3] game (2003) was one of the first attempts to expand the boundaries of a game. Intriguing messages (SMS, emails) could be sent by the game server, any time regardless of the player connection. The line between the game and daily life (the border of the magic circle) became uncertain. Human Pacman[4] (2004) breathed life back into the arcade game from the 80's, by replacing the mazes and characters with real urban areas and humans. Each player had a mobile phone and was geo-localized. The European project Iperg[5] (Integrated Project one Pervasive Gaming, 2004-08) offered three different toolkits (augmented reality, mobile telephony and ubicomp), making it possible to deploy a pervasive game. The project also originated several pervasive games developed with these tools. MYHT[6] (2008) took up the idea of the city as playground: around ten players find their twin partner, according to their cardiac rhythm. Each player was equipped with a mobile phone and a heart rate sensor communicating via Bluetooth. In other works, one can see technologies with shorter range like Bluetooth or zigbee: either the environment is more constrained (interior only), or the requirement is only the replacement of a wired connection (point-to-point), Liu and Ma[7] or Borriello and al.[8]. Overall, the popular networking technologies for the development of pervasive games remain Wifi and GSM, for covering indoor and outdoor environments.

In the following sections, we present our experiment in which a zigbee mesh was deployed to monitor the players in real-time and thus to provide in real-time information to the game engine.

3. UBIQUITOUS NETWORKING

Although there are many options for wireless technology (see Table 1), player mobility necessitates a flexible wireless technology involving roaming and routing. RF and Bluetooth are more dedicated to peer to peer (*ad hoc*) networks and were therefore bad choices. GSM, WiFi and ZB remained as candidates. ZB was the best choice, for cost, and power consumption, as well as other factors not mentioned in Table 1: hardware size, electronic integration and ease of use.

We have two types of zigbee modules, manufactured by Digi (formerly Maxstream): XBee series 1 (XBs1) and XBee series 2 (XBs2). The low electricity consumption of this equipment makes it an ideal candidate for embedded applications. The main difference the two models lie in the capacity of XBs2 to provide multi-hops routing capabilities, XBs1 offering only short range point-to-point or broadcast connections.

Table 1. Wireless devices comparison

Type	Range	Data Rate	Cost	Energy	Interface
GSM GM862 Cellular Quad Band Module	Like a mobile phone	57.6kbits/s	100$	17-250mA	Serial
WiFi WiFly GSX Serial Module - Roving Networks	100m	2.7 Mbits/s	70$	40-210mA	Serial
ZigBee ZigBee/ZNet2.5 Module	120m	250 kbits/s	38$	35-38mA	Serial
RF RF Link Transmitter - 434MHz	120m	Few kbits/s	4$	3.5mA	Serial
Bluetooth Bluetooth Module - Roving Networks	10m	1.5 Mbits/s	60$	30mA	Serial

Table 1 does not include Bluetooth 4.0 (BLE) since this project ended before BLE became available. Basically, it shows that ZB stands on an average position for each criterion.

Although the final objective was to deploy a multi-hops WSN, the XBs1 modules allow communication between the sensors and a remote server, and to test the emission frequencies. Thus the messages are transmitted (timestamp; ECGval; AXxVal; AXyVal; AXzVal) without any loss by sampling at 500Hz and by transmitting at 57600 bauds. This configuration incurs a significant degree of loss and duplicate messages. In order to preserve a sufficient sampling rate for the treatment of signals (in particular for the ECG which requires a sampling rate around 400Hz), we embed the signal processing and their interpretation on the mobile device (i.e. on the microcontroller). After some optimizations, we lowered the message frequency to 1Hz which is sufficient to send the movement and heart rate. With this frequency, all congestion and message loss was removed.

We created the embedded device (built upon an ATMEGA160AU microcontroller) ourselves, as well as the 14 ZB routers to cover the 3 floors of the museum and the ZB-USB converter on the server. The routers are autonomous and are supplied by a rechargeable 9V battery.

The ZB network is self-healing: each new ZB module is auto-configured. This functionality was important for two reasons. First, there was no need to link the configuration of a particular node depending on its location. Each deployed node, once powered, broadcasts its IDs so that the closer nodes (i.e. within the transmission range) are aware of its existence, and then propagates, hopping the entire network. Second, this feature preserves an uninterrupted transmission during player movement (roaming), since the "mobile node" is always aware of its surrounding nodes.

4. RESULTS

Over the duration of a game session - which could range from 30mn to 1.5 hours, depending of the player skill - we were able to monitor two players simultaneously. The amount of lost data was under 2%. This good score allowed us to maintain the data stream to the game engine without any significant perturbation. For instance, the heart rate signal never had any rapid significant variations, allowing the loss of a small amount of data without causing disturbance. Although some node power dropped, the WSN compensated for this and the data stream was never significantly interrupted.

5. CONCLUSION

In our project we continuously and simultaneously monitored the heart rate and acceleration of two players moving freely within the space of a building comprising 3 floors, each measuring 2000m². The WSN, built around XBs2 modules, made it possible to cover all rooms of the museum, while respecting the constraints imposed by the institution: discretion of the installation, respect of the building and robustness. The 9V battery supplying some nodes had 5 hours of autonomy, enough for a 1.5 hour gaming session. This hardware architecture presents some excellent features. First, the deployed network is meshed. Hence we can rely on a minimum quality of service (i.e. routing and switching to alternative routes in case of a node breaking down. Second, this technology can be either wall powered or battery powered, avoiding the need to install mains power lines in locations that do not already have them. Third, the shape of a node is very tiny (the size of the XBee module is 24 x 28 x 9 mm) allowing it to be placed in really different places (in the ceiling, stuck to a wall, on furniture, on the floor, etc.)

We are currently working on optimizing the number of routers needed to cover the museum and also on improving the integration on the gaming device (temperature and galvanic skin response ones) so that the issues discussed earlier can be compensated for. One more perspective remains in our ability to use this ZB network to accurately geo-localize people in the museum in order to enhance the game experience by sending contextualised content to the players.

ACKNOWLEDGMENTS

We rise to honour and pay homage to a special person departed from us too soon. His sense of humour, spirit of camaraderie, good-natured attitude and counsel are sincerely missed. Though Alex is profoundly missed, his contributions to search will prevail. This article is dedicated to him.

We would like to thank all the members of the PLUG project, especially MATLOG (http://www.matlog.com/) for their patience and help during the hardware orderings. The PLUG project was funded by the French National Agency for Research (ANR).

REFERENCES

1. Dupire, J., Topol, A., Froment, S., Ginibre, M., Mader, S., Sarafian, A., Schwartz, A. et D. Soriano, D. 2009. AZ66 . In IEEE ICE-GIC'09, IEEE Consumer Electronics Society's Games Innovations Conference, London, August, 2009;
2. Dupire, J., Gal, V. et Topol, A. 2008. Physiological Player Sensing: New Interaction Devices for Video Games . In ICEC'08, IFIP 7th Int. Conf. on Entertainment Computing, Pittsburgh, September, LNCS 5309, Springer Verlag.
3. http://www.inmemoriam-thegame.com/
4. Cheok, A., Goh, K. H., Liu, W., Farbiz, F., Fong, S. W., Teo, S. L., Li, Y., and Yang, X. 2004. Human Pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. Personal Ubiquitous Comput. 8, 2 (May. 2004).
5. Lindt, I., Ohlenburg, J., Pankoke-Babatz, U., Prinz, W., and Ghellal, S. 2006. Combining multiple gaming interfaces in epidemic menace. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada, April 22 - 27, 2006). CHI '06. ACM, New York, NY, 213-218.
6. http://xilabs.fr/2009/05/meet-your-heartbeat-twin/
7. Borriello, G., Hartung, C., Hemingway, B., Koscher, K., and Mayton, B. 2008. Multi-player soccer and wireless embedded systems. In Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education (Portland, OR, USA, March 12 - 15, 2008). SIGCSE '08. ACM, New York, NY, 82-86.
8. Liu, L. and Ma, H. 2006. Wireless sensor network based mobile pet game. In Proceedings of 5th ACM SIGCOMM Workshop on Network and System Support For Games (Singapore, October 30 - 31, 2006). NetGames '06. ACM, New York, NY, 30.
9. Plug Secrets of The Museum - Trailer -http://www.dailymotion.com/video/xalc8t_plug-secrets-of-the-museum-trailer_creation#from=embediframe

ISBN: 978-989-8533-38-8 © 2015

THE DEVELOPMENT OF A SENSORY GAME FOR ELDERLY PEOPLE UTILIZING PHYSICAL CONTACT

Shinnosuke Ide, Yumu Horii, Hisashi Sato and Takashi Nakamura
Kanagawa Institute of Technology, 1030 Shimoogino Atsugi Kanagawa Japan

ABSTRACT

Recently, the shortage of care personnel is a problem in Japan. Rehabilitation, which is necessary for the elderly to live without nursing care, is a monotone activity that the elderly often feels stress to carry on. In this study, we noticed the benefit of the movement of playing tag to touch bodies, and developed a game for rehabilitation by adopting this movement. We developed (i) a unique game controller easy to handle by the elderly and consisting of an apron and an electrically conductive fabric, with which the judgment as to whether a body of a player is touched or not can be made, and (ii) a game system in which the body movement in gymnastic exercises for rehabilitation is adopted.
As a result of actual use of the game by the elderly, the game was evaluated such that the game per se was mostly a good rehabilitation game, but in conclusion, it was considered as being too difficult for the elderly because of its fast tempo.

KEYWORDS

Rehabilitation, Game for elder people, Body-contacting game, electro conductive fabric

1. INTRODUCTION

Recently, the lack of sufficient caregivers has been posing a serious problem in Japanese society. The problem is mainly attributable to the relative decrease of caregiving personnel in the increasingly aging society. Other reasons may be the growing elderly population living alone and understaffed nursing homes due to professional caregivers leaving their jobs[1].

All these causes are social problems and require substantial time to solve the situations. Under such circumstances, rehabilitation methods that can be continuously used to encourage elderly people to look after themselves without the caregivers' supports are necessary. However, as tedious rehabilitation practices that require elderly people with decreased physical strength to continue monotonous movements often give them stress and the feeling of boredom, professional caregivers generally accompany them in the actual rehabilitation exercises.

In order to solve this dilemma, sensory games to complement rehabilitation practices have been studied. A sensory game is a game that requires the player to make big body movements. In addition to the sense of purpose and reduction of monotonous feeling associated with rehabilitations by presenting exercises as a game, these programs aim to provide adequate rehabilitation exercises in the absence of caregivers by mechanically giving instructions to players and managing data. The downside of such rehabilitation games is that most of them are large and expensive. Because their availability is limited to spacious places like nursing homes, it is difficult to install them in individual households. They are therefore not easily accessible for housebound patients living alone.

Elderly people are not familiar with digital games, and their physical and cognitive functions are declined. Due to these aspects, they are likely to get confused when they are asked to handle complicated operations or given too much information[2]. These features require the presence of an operating device which elderly people find comfortable to handle.

This study aims to develop a small-size rehabilitation game system designed for homebound patients unable to use large-size rehabilitation games. It also creates an operating device that elderly people can use without difficulty in their rehabilitation exercises.

2. CURRENT TASKS

In addition to the aging, the elderly people's physical and cognitive functions decline as their opportunities to move their bodies reduce due to the lack of work, etc. In this study, the points of improvement concerning the declined abilities of elderly people have been selected in accordance with "the Change in the Elderly People's Physical Abilities and Their Behavioral Characteristics"[3]. One of the features that significantly differentiate the game in this study from games in general is that it targets elderly people, who are not familiar with games. In particular, their deteriorated functions must be taken into account in designing the game aiming to improve the players' physical and cognitive functions, which is the purpose of this study.

Also, elderly people's moods are easy to swing, characterized by their susceptibility to stress with just a little obstacle. If the rehabilitation game involves comparison with other players or other disturbing elements typical for games, the player might experience increasing stress and reject to use it continuously. Because rehabilitation practices are meaningless unless the patient continues them, it is essential for the game to avoid the risk of being abandoned in the middle of the rehabilitation program.

As a sensory game requires its player to make big body movements, it has to be able to detect the movements. At the moment, three mainstream methods to detect players' movements are infrared sensor, camera photographing, and the use of special controller.

Because the infrared sensor and camera both require certain space and are easily affected by the environments, a special controller was developed and used for this study.

3. OUTLINE OF THE PROPOSED SYSTEM

In view of the issues raised in the preceding chapter, this study aims to develop a game that is operable with as simple actions as possible. In the game developed this time, movement instructions are given in picture forms so as to easily convey intentions to elderly players and enable them to engage in rehabilitative movements simply by copying the images without going through complex mental processes.

At creating a rehabilitation game for this study, we established a hypothesis that the operation utilizing body touches may increase the game's operability. The reason behind the hypothesis is action rhymes written for small children.

Compared to dances, action rhymes involve simple movements and are easy to remember for children who are not familiar with the lyrics. As typically represented in Yankee Doodle (*arupusu ichimanjaku*) or Bingo, the performers are frequently required to touch their bodies by clapping hands, tapping knees, or putting hands on waists or shoulders. From these facts, we assumed that actions involving body touches are easy to remember and may improve the game's operability.

For this study, we referred to a rehabilitation exercise[4] to improve the players' physical ability and focused on reproducing the movement designed to improve the lower back function. The lower back exercise was chosen because the part is the center of the whole body and yet often suffers fissures and wearing of the spine and sacrum with age. In order to lessen the burden on the weakened bones, the back muscle and other muscles surrounding the bones must be strengthened. The game incorporates twisting and bending movements of the body to train muscles surrounding the lower back to reduce the load on the bones.

For the controller of the rehabilitation game, we decided to develop an original controller that directly detects contacts through sensors attached to certain body parts, as it was deemed difficult to detect contacts with external sensors. The controller must be easy to put on, as an elderly player may find it too stressful to put on a troublesome gear. Taking these factors into consideration, we decided to develop a controller utilizing an apron that is easy to put on and covers shoulder, waist, and knees, where sensors are attached for the rehabilitation purpose. The figure below shows the controller developed for this study (Figure 1).

As the sensors attached to the apron is designed to touch the devices attached to the player's body, some players would find it painful when they press the body parts attached with sensors due to the extra pressure added by the devices. To avoid this, we used an electro conductive fabric this time.

In this study, an electro conductive fabric was used as the sensor. An electro conductive fabric is a cloth made of metal-coated nylon threads. While retaining the conductivity, it can be cut and reshaped just as an ordinary cloth. As the rehabilitation game used in this study did not require complicated input and output controls, the electro conductive fabric was sewn to where the sensors were attached to detect the input and output through the presence of energization as an alternative to mechanical sensors.

Figure 1. The Game Controlor

As the electro conductive fabric did not cause irregularity on the surface where sensors were attached, the players did not feel too strong pressure when they touch these parts. Furthermore, as the sensors detect the touches directly through energization without the need for switches, the time difference between a touch and detection became almost nonexistent. As for the durability, it was concluded that there would be no problem unless the device suffers an unexpected level of pressure.

The electro conductive fabrics sewn to the sensor parts covered extensive areas as a whole to accommodate players with different physical constitutions. In addition, each sensor part was presented with a color-coded frame to help players distinguish different body parts not only by names but also by colors.

The controller used in this study was a plain, ordinary apron attached with sensors covering areas of both sides of the chest and waist, and knees. The backside of the apron was attached with a microcomputer called Arduino. Each sensor part was connected to Arduino through copper wires. The connected Arduino was then connected through USB to a PC, which controlled the operation.

The flow of the game created for this study is basically as follows: first, a character appears on the screen. The player touches the controller so as to stand like a mirror reflection of the character. The player then mirror-copy the character for a certain period of length and get the score. In this rehabilitation game, time limit is the only obstacle, as the program tries to keep obstacles as minimal as possible. Also, a human-shaped action model is featured in the screen to enable the player to instinctively understand what to do, as instructions given in letters would require the player longer time to digest, and, when the player fails in an attempt, resuming the correct position would also take long (Figure 2).

Figure 2. Game Play of the Proposed Game

The game has two modes; one is the tutorial mode, and the other is the game mode. The tutorial mode is designed to let the player get used to the operation, but it also intends to help them warm up through exercises equally given to all body parts. As the game mode gives out instructions randomly, the program expects the player to go through the tutorial mode before moving on to the game mode. As the game mode does not give any hint as to the next move until the instruction is given, the player must reflexively respond and touch the correct parts of the body in order to get a high score.

The game has two modes; one is the tutorial mode, and the other is the game mode. The tutorial mode is designed to let the player get used to the operation, but it also intends to help them warm up through exercises equally given to all body parts. As the game mode gives out instructions randomly, the program expects the player to go through the tutorial mode before moving on to the game mode. As the game mode does not give any hint as to the next move until the instruction is given, the player must reflexively respond and touch the correct parts of the body in order to get a high score.

This feature encourages improved reflexes in addition to motor ability. In order to avoid giving repetitive instructions, the program also remembers the previous instructions and re-selects a new move if the same instruction is consecutively generated or the same type of movement is repeated based on random numbers.

Japanese summer festivals inspire the general design of the game. We also picked calm, Japanese-tone BGM to accompany the game. These choices were made based on an assumption that the elderly players may find a Japanese design familiar. Besides, the movement instructions used in the game are similar to the movements used in the traditional bon dances, making them easier to be incorporated in the summer festival design of the instruction model.

4. EXPERIMENT

The aim of the experiment is to verify whether the problems presented in Chapter 2 is solved. To do so, we actually had elderly people at the age of 70 and above play the rehabilitation game we created and verified through a questionnaire whether it solved the problems. Because the participation of the elderly people was necessary to conduct the experiment, we asked for the cooperation of both nursing homes and elderly people living alone. The contents of the questionnaire are as follows:

(a) Was it fun to play the game?
(b) Were you exhausted by playing the game?
(c) Did you find the game difficult?
(d) Was it difficult to put on the controller?
(e) Do you want to play the game again?
(f) Which do you prefer between the ordinary rehabilitation exercises and the game?

The participants were asked to answer the above by either “yes” or “no” and also provide optional comments. These items, along with the comments, were then used for the evaluation of the game based on the participants’ game scores, ages, and gender.

The questionnaire results are as follows (Table 1).

Table 1. Questionnaire Results

Questions	Yes	No
(a)	9	1
(b)	8	2
(c)	10	0
(d)	2	8
(e)	9	1
(f)	7	3

5. DISCUSSIONS

Because the players were able to follow the actions expected in the program, the rehabilitative function of the game was proved effective. On the other hand, the game also incorporated the training for reflex movements and demanded the players to follow a quick flow of the game. As a result, the program kept on rushing to the next movements, making each move momentary. As for the movements made by the players, many participants managed to make big movements, but their energy consumption within a given time frame became intensified as well. Hence, the movements of many participants slackened towards the end of the set time frame.

This feature as a result might have reduced the intended effect on the improved motor ability. As a point of improvement, we would like to include instructions asking the player to press a specific body part for a certain length of time while maintaining instructions equivalent to the existing program.

Although we intended to prioritize simple operation to make it friendly to elderly people, the questionnaire result revealed that all participants found it difficult. Through observations of the participants actually playing the game, it was clarified that they found the game difficult when they got confused in their attempt to quickly follow the instructions whether or not they aimed for high scores. For many participants, the confusion particularly increased when they touched a wrong place and hastily attempted to touch the correct position. Nevertheless, most of the participants laughed and looked excited while playing, including the moments of hasty attempts. In the questionnaire, many of them responded that the game was fun. Since the obstacle incorporated in the game is the element to trigger fun and aspiration for improvement, the game was successful in that sense. At the same time, a few participants seemed to find the moments of confusion stressful. The game requires improvements so that more people can find it friendly. As mentioned in the previous section, one suggested reason for the confusion is the quick speed of the game.

One way to improve the issue is to display the total number of movements or something else in between two instructions immediately after the start of the game. By creating intervals in this way, the players may find it less confusing. By gradually shortening the interval display every time the player completes a certain number of movements afterwards, the game should maintain the original tempo for those who are used to playing the game.

Although some people looked confused to put on an apron for the first time, all participants managed to put on the gear on their own. Similarly, most of them quickly understood which part of the gear should go where based on the color codes and their locations. As the participants accepted the appearance of the gear without being taken aback, the apron was a success as a sensory controller designed to touch the player's body. However, improved operability was not suggested in the result.

Also, some players touched areas slightly apart from the instructed positions due to their physical constitutions, such as their sitting heights or arm lengths. While there was no incident of failed detection of players' response this time, there remains a possibility that some players touch outside the sensor area. One way of improvement may be to make the sensors detachable so that each player can move the sensor locations to where they are easy to touch.

6. CONCLUSION

We developed a rehabilitative game designed for elderly people that involves body touch movements. Upon developing a game system friendly to elderly people, we evaluated its effect as a rehabilitation program incorporating rehabilitative movements. Body contact was used for the evaluation of operability and the participants' improved understanding. The completed game was then put to actual use by elderly people and submitted for their appraisal. The apron-shaped controller was easy to put on, and the simple game contents were well received by the elderly participants who are not very familiar with games. Although the use of body contacts was useful in reproducing the movements used in the rehabilitative exercise, the expected operability improvement through the use of specific body parts was not identified. As points of improvement for the future development, we suggest the introduction of detachable sensors to improve operability of the controller, and for the game system, the inclusion of intervals between instructions can be suggested as a method to adjust the game speed.

REFERENCES

[1] Ministry of Health, Labour and Welfare in Japan, http://www.mhlw.go.jp/seisaku/09.html .

[2] Matsuguma H. Et.al, 2012, Significance of serious game for rehabilitation in super-aging society (in Japanese), *The Journal of Information Science and Technology Association*, Vol. 62 NO. 12, pp 520-526.

[3] Naka Yoshino Fire Department, http://www.nakayoshino.or.jp/joho_mienaihonoo.htm .

[4] Kumamoto Rehabilitation Hospital, http://www.marutakai.or.jp/kumareha/kcr/htlm/kikucyan%20isu.htm.

HEALTH GAMES

Fabiane Barili and Marta Bez
Universidade Feevale

ABSTRACT

This article presents an introduction to the main concepts of games discussing its application in the process of teaching and learning in several areas. The focus will be given on games in the virtual patient simulators type format, to facilitate the process of learning the clinical reasoning by students in the health area. Lastly, it will be presented the Health Simulator project, which is currently being developed in a Brazilian university.

KEYWORDS

Health Games, Serious Games, Gamification, Simulators.

1. INTRODUCTION

Games have several definitions that have been studied by some authors. According to the study by Alves (2007), Huizinga (2001) defines them as:

> "The game is more than a physiological phenomenon or a psychological reflection. It exceeds the limits of purely physical activity or biological. It is a significant function that contains a certain direction. In the game there is something 'in game' that transcends the immediate needs of life and gives meaning to the action. Every game means something. "(Huizinga, 2001, p. 4)

Alves approaches the perspective of Huizinga (2007, p.22) as the games being an intellectual technology, an element that organizes and modifies the cognitive ecology, which promotes the construction or reorganization of cognitive functions with the memory, attention, creativity and imagination. The digital games are an evolution of the traditional games, they are graphically represented by a software that is responsible for carrying out the simulation of the game matches. According to Shuytema (2008, p. 7) the digital games are playful activities that have a series of actions and decisions, where there are rules in the game's universe, that result in a final condition. These rules and the universe created within the digital game exist to provide a structure and context for the actions of the game, furthermore, situations are created to challenge the player and the richness of the context, the challenge, the emotion connection and the fun involved in the game are what determine the success of it.

As games have a definition, there are different classifications differentiating the characteristics that they present. According to Azevedo (2005) games can be categorized by gender, for example, adventurous games are based on stories that focus on puzzle solving, however, action games are more realistic in time, usually represented by shooting strategies, among other genres. As stated, games have individual characteristics that categorize them and can be differentiated. This article will specifically refer to games characterized by being an aid tool for general people and also training and learning activities for specific group. This sort of games may be considered Serious Games once it is defined as it follows:

> "A game is considered serious when is applied as a simulation in order to enable a computing environment using the training of professionals or as an aid tool to form the student learning." (SUSI, 2007)

This article has as main focus the Serious Games and their results. This article focuses in Serious Games and their effective results. Games which help positively will be analyzed, specially their effects in the medical area, where games may be used as training tools by medical students.

For the projects that were developed to become games, it is necessary to talk about the study of the concepts as the gamification and its characteristics, in which it influences directly in activity projects and its results.

2. GAMIFICATION

The term gamification has a consensus among some theorists. Kim (2010) defines as "the use of the techniques of the games to make people more engaged and make funnier activities", Kapp (2012) defines it as "game using – mechanical base, aesthetics approach and game thinking to involve people, motivate actions, promote learning and solve problems " as well as Deterding (2011) who says that "the use of the game design elements in a non-game context." As it can be seen, gamification is about using elements and mechanics of games to promote learning, problem solving, among other applications.

Gamification is currently used in several areas, for instance in products marketing, teaching and learning, motivational activities and even as a tool to help change the perspective of some activities, as it is known that games are funnier and more engaging to users.

An example of gamification that was used in learning is the Duolinguo (2013), which consists in an application that teaches and helps people to practice other languages. There are several challenges to be overcome, including hearts that the user can't miss, the "linguotes" he wins if he helps some colleague in a difficult situation, besides competitions with other users or with the robot programmed within the application. In terms of education, the focus is on the translation of words or phrases from the language of the student and the target language. It also includes exercises where the student listens and must transcribe in the dictation language. In short, the user can compete and have fun while learning.

3. GAMES DEVELOPED IN THE HEALTH AREA

Some universities have already developed the Serious Games aiming to create a real simulation to develop objective in learning and assistance for the professionals in the health area.

Projects developed as the SimDecs (2012) that consist in a simulator that has a mechanism to simulate calls and daily medical cases. Furthermore, they have a mechanism of artificial intelligence, which evaluates the decision-making of medical students, in this case, the SimDecs players. The simulator works as a database where the clinical cases that will be presented to students who must find the solution will be recorded. The coordinators of the simulation can add several cases to expand the clinical cases library. Students are rigorously evaluated in several forms, reliability, trust, unnecessary tests that were ordered to the patient in the clinical case to be solved. In the case of SimDecs, there are numerous benefits provided by the simulation, such as helping the students understand some complex relations, that would require expensive equipment, avoid risks to patients and students, allow the students to develop other methods and strategies for solving the problem (Bez, 2013).

Researchers at the University of Iowa developed the GeriaSims (2008). It consists of a group of virtual patients whose the focus is on the geriatrics and on the nature of the care needed for this type of case. This is available to medical students, residents in the areas of the primary care and doctors. An example of GeriaSims is the case of a patient with Alzheimer, who can be monitored from the start of the diagnosis until his death. Through a registration system implemented by GeriaSims, it was possible to note that the simulator was adopted by several medical training programs (Orthon and Mulheasen, 2008).

The Web-SP is a simulator of virtual patients developed by the Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Sweden (Botuzatu, 2010). This simulator basically works with the resolution of virtual patient cases, the students can make interviews, physical and additional exams, in order to get a diagnosis and indicate a treatment. The College of Medicine of the University of Bosque, in Bogotá, adopted this simulator and the cases used were created from real records of patients of the university hospitals. Thus, these cases contained photographs of patients and means of relative diagnosis.

The simulators analyzed SimDecs, GeriaSims and Web-SP demonstrate in practice the advantages of learning as cost-benefit in cases of learning in the medical area. Several universities have adopted this measure to ensure the safety of patients and control the costs involved in the pratical learning process of the medical students, but without interfering in order for students to learn the most. The results obtained from each simulation consist of being able to evaluate the decision-making of medical students, providing them thought of different alternatives to be adopted in time to solve a clinical case, among others.

4. HEALTH SIMULATOR

The Health Simulator is a project that is being developed at the University Feevale in partnership with teachers and students from several courses at this University. This project consists of a simulator of clinical cases of Virtual Patient type, which can support the teaching and learning processes in the health area. The simulator was developed in 3D graphics with realistic art style, characters of several ethnics and a variety of scenarios based locally where the medical care takes place. The project in development is divided in two distinct parts: back-end and front-end.

4.1 Front-End

The process of front-end refers to the part of the simulator for the students of the Digital Games course, which will be in the form of a Serious Game, with scenarios and characters that represent the entire environment and real medical care.

For the Health, there will be four categories of characters: Doctors, Patients, Nurses and Dentists, all specified as men and women. The characters are also divided by ethnics, being brown, white, black, Indian and yellow. These characters have different ages and weights as thin, ideal weight or obese. A study about the characteristics of health professionals was realized and because of this it was developed three-dimensional models. In the end, there are approximately 260 character models just for the category of patients.

The scenarios have produced a commutative system, in other words, the models produced are reused in several different scenarios. They are classified as offices or hospitals of A, B or C class, and also a hospital of the Unified Health System (SUS). In the production of scenarios, documents were made with planks of moodboard (Figure 1) and spreadsheets containing the main items of the scenarios.

Figure 1. Mood board office class C

All the elements of the scenarios and the characters were developed and modeled in 3D, using 3D Studio software Max. Both the characters and the scenarios developed are incorporated into the game engine, in this case the Unity 3D was used (Figure 2).

ISBN: 978-989-8533-38-8 © 2015

Figure 2. Office class C

4.2 Back-End

The back-end process is divided in a few steps: Creation of knowledge; Administration interface and Web Service communication.

The stage of knowledge modeling consists in delimiting the knowledge that will be represented by the specialist. Thus, it makes use of a clinic directrix, a systematic form to orient and delimitate the content that will be develop from evidences. For this, it is necessary to develop a statistical model of knowledge representation, with a Bayesian network (BN). This network is the base of knowledge to create the clinical case later.

The management interface is a web interface (Figure 3) where the teacher has access to the signs and symptoms present on the network, possible previous history of the patient as well as physical tests and additional exams. Clinical cases are modeled through the system that presents the outcome of each clinical case when it is selected, all being stored in a database (DB). A clinical case generated in the Health Simulator is stored and stay available for use by students in the game.

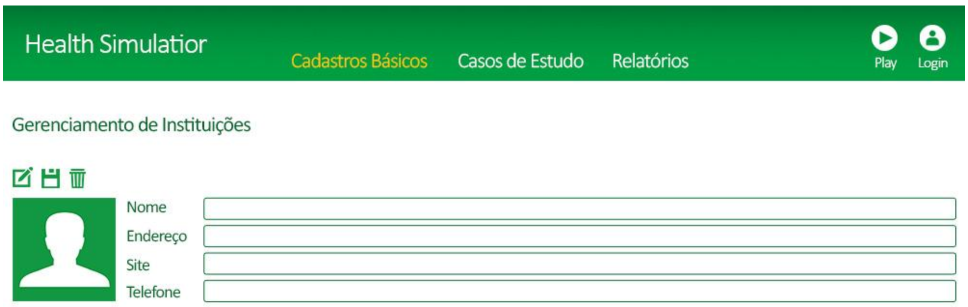


Figure 3. Web Interface

The Web Service Communications allows a greater compatibility of software developed in different platforms. This way, the Web Services give a solution that sets the standard patter for the exchange of messages between the client's applications (game/client) and a service provider (back-end).

5. CONCLUSION

The use of simulators as teaching and learning alternatives has proved an excellent choice to bring the medical students closer to the reality they will find in hospitals and clinics. Even though there are many simulators, as presented, there is the necessity for a more realistic simulator, allowing the student to have more flexibility to test their diagnostic hypotheses.

Health Simulator made the simulation more real in comparison to other presented simulators. In this simulator countless benefits foucused on student learning may be found, such as: cost and risk reduction regarding complex experiences; scientific knowledge application; search for alternative methods and strategies when solving the same study case; acquired knowledge reiforcement; among others.

Analyzing the advantages and examples of the simulators presented, there is a possibility to use some of seen techniques to produce other simulators, as the character and scenery development technique, which aims the project speed. For instance, the characters were created in a way that models could be reutilized and the scenery elements were developed in order to compose several different sceneries.

The game is being developed at Feevale University with guidance of Doctor Marta Bez, alongside with the collaboration of scholars, volunteers and Computer Science and Design teachers. At the moment, Health Simulator has three concluded scenaries: Class C Office, Public Unified Health System Hospital (SUS Hospital) and hospital bed. The character doctors are concluded and the patients are being developed. The project is planned to be firstly applied in Brazil, however it is expected to be extended to international universities as well.

REFERENCES

ALVES, L. *Game Over: Jogos Eletrônicos e Violênca.* São Paulo: Futura, 2005.

AZEVEDO, E. *Desenvolvimento de Jogos 3D e Aplicações em Realidade Virtual.* Rio de Janeiro: Campus, 2005.

BARROS, P. R. M.; CAZELLA, S. C.; BEZ, M. R.; FLORES, C. D.; DHAMER, A.; MOSSMANN, J. B.; FONSECA, J.; MARONI, V. (2012) *Um Simulador de Casos Clínicos Complexos no Processo de Aprendizagem em Saúde.* REMOTE. Revista Novas Tecnologias na Educação, v. 12, p. 1-11.

BEZ, M. R. (2013) *Construção de um Modelo para o Uso de Simuladores na Implementação de Métodos Ativos de Aprendizagem das Escolas de Medicina*. Porto Alegre, 2013. 314 f. Tese (Doutorado em Informática na Educação) – PGIE. CINTED – Universidade Federal do Rio Grande do Sul, Porto Alegre.

BEZ, M. R.; FLORES, C. D.; MOSSMANN, J. B.; BRANCO, M.; MENDES, T.; (2013) *SimDecs: Um Simulador Baseado em Técnicas de IA,* Novo Hamburgo, 2013 – Universidade Feevale, Novo Hamburgo – Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Rio grande do Sul.

HUIZINGA, J. Homo Ludens: *o jogo como elemento da cultura*. São Paulo: Perspectiva, 1980.

MCGONIGAL, J. *Realidade em jogo: por que os games nos tornam melhores e como eles podem mudar o mundo*. Rio de Janeiro: Best Seller, 2012.

NAVARRO, G. (2013) *Gamificação: a transformaçãoo do conceito termo jogo no contexto pós modernidade*. São Paulo – CELACC/ECA - Universidade de São Paulo, São Paulo.

PAZ, T.; FUENTES, L.; NEVES, I. B; LEMOS, V.; ALVES, L. (2014) *Dispositivos móveis e gamificação: interfaces lúdicas em novas práticas educativas.* Salvador - Departamento de Educação – Universidade do Estado da Bahia, Salvador.

ISBN: 978-989-8533-38-8 © 2015

AN EFFICIENT METHOD FOR DESIGNING LEARNING GAMES FOR MATHEMATICS

Vani Kalloo[1], Permanand Mohan[1] and Kinshuk[2]
[1]*The University of the West Indies, Trinidad & Tobago*
[2]*School of Computing and Information Systems, Athabasca University, Canada*

ABSTRACT

Researchers have shown that games can be extremely effective learning tools; however, designing successful mathematics learning games is challenging. In this paper, we present a unique method for designing games for learning mathematics, in an effort to improve the traditional process by ensuring that there is sufficient integration of mathematics content and that it is more systematic and guided. This method for creating mathematics learning games is based on the study of several factors such as game styles, the mathematics curriculum and past research experiences in mathematics learning games. The evaluation design of the method is presented. This method is improved compared to the traditional design process because it guides the designer from a general perspective, but still gives room for being creative.

KEYWORDS

Learning Games, Game Design, Reuse, Game Patterns, Educational Games, Instructional Design.

1. INTRODUCTION

Players are captivated by games, while learners are repelled by mathematics. The high failure rates of mathematics throughout the Caribbean verifies this (Caribexams, 2008). Many researchers agree that games are extremely appealing, motivating, engaging, intriguing, fascinating, interesting (Garris et al. 2002; Connolly et al. 2007) and suitable for learning mathematics (Koster, 2013; Gee, 2003). Therefore, games can be an effective tool for learning mathematics. Conversely, not all games created are motivating and appealing. This leads to the problem that it is difficult to design successful and appealing mathematics learning games. A contributing factor which can make design difficult is ensuring that the mathematics content is an integral part of the game play. Games such as Pokemon (Lin, 2007) or Thinklets (Jonker & Wijers, 2008), are good examples of integrating the mathematics content into the games. The problem tackled in this paper is that it is difficult to design games that appeal to the players, integrate mathematics content directly into the game and successfully help the players improve their mathematics skills. In this paper, we present a reuse method which can be used as a guide to designing mathematics learning games in an effort to make the process more systematic.

2. LITERATURE REVIEW

Several researchers have identified that when fun and appeal from games are translated into learning, great gains in education can be attained. Many researchers have obtained favorable results showing that games are valuable for learning (Ke & Grabowski, 2007; Gros, 2007). Games are difficult to design (Rogers, 2010) and especially so for mathematics games. Several researchers have attempted to address this problem. Torrente et al. (2008) tackled the problem by proposing instructor-oriented authoring tools. These tools focus on the abstraction of programming tasks and support aspects specific to the educational domain. Similarly, Abdullah et al. (2008) proposed a simple authoring tool to help teachers produce and personalize their own game-based educational content. Gómez-Martín et al. (2007) proposed a two level data driven architecture where domain experts create the exercises and game designers create a game template. The problem with this approach is that the games and the exercises may not be well-integrated. Some researchers suggested that *game mods* could provide a means for educators to use the quality and the basic format of commercial games to create customizable instructions for students (Fanning, 2006). *Game mod* is the modification of an existing commercial computer-based game. However, the problem still exists because it is difficult to get teachers to use such technology. Our research explores an aspect of game-based learning by looking at improving the design of learning games for mathematics.

3. METHOD FOR DEVELOPING MATHEMATICS LEARNING GAMES

Our method for developing mathematics learning games is a process which guides the designer through mapping a mathematics topic or skill to game styles, in an effort to help them design a learning game. A game style or type is defined as any game approach describing the general game play for a specific game, for example puzzle solving, collecting and merging. Prensky (2001) identified several different learning content and possible game styles to suit each. However, Prensky's suggestions are based on general learning content, while our research is specifically focused on learning mathematics and includes a guided process for designing the mathematics learning game.

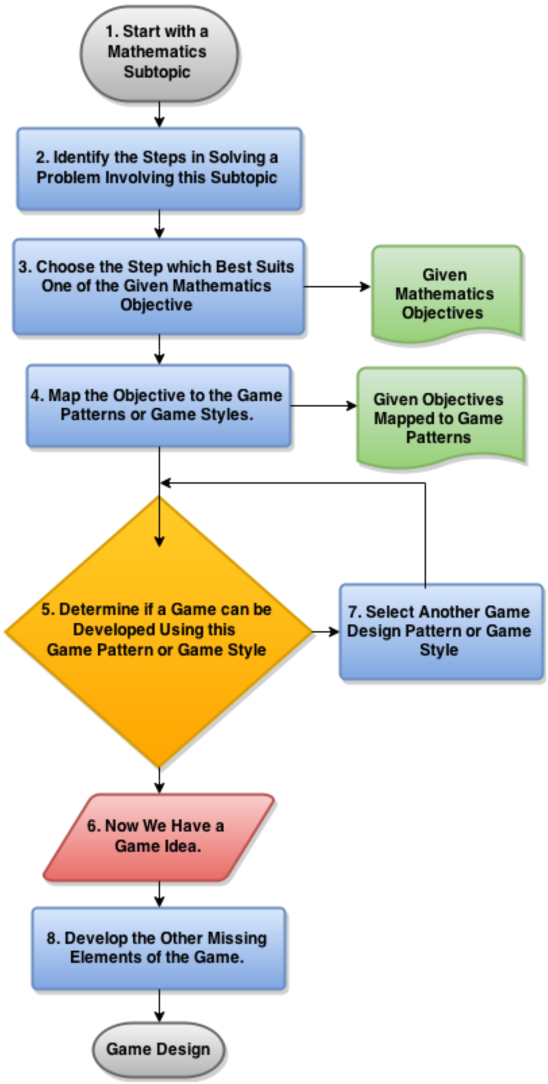


Figure 1. Method for Developing Games

Figure 1 illustrates our method for developing mathematics learning games. The method consists of a set of steps which guides the designer through the game design process. The method involves selecting a mathematics topic and mapping it to a suitable game style. After mapping the mathematics topic to the game style, the designer now has the main idea for the game. He or she then has to work out the other details of the game guided by specific principles in order to ensure that the game is recognized as a game and the game play involves the mathematics topic chosen. These principles are those which results in the player being captivated by the game, such as the ability to attract and sustain the players' attention, motivate, engage and interest the player.

Step 1 guides the designer to select a mathematics topic for the focus of the game. Depending on the complexity of the mathematics topic, it may be best to break topics into smaller components which we refer to as subtopics. Therefore, the game designer can break down this topic into as small a topic as he or she wishes. For example, if solving linear equations are chosen, then this can be broken down into changing the subject of the formula, simplification of algebraic terms, factorization, division, multiplication, addition and subtraction.

Step 2 requires the game designer to identify each part of the process of solving a problem based on the selected subtopic. It is expected that the designer will get a better idea of the topic when the problem and its solution are examined. For instance, for the subtopic simplification of algebraic terms, the procedure for solving a problem is categorizing alike terms together and then adding or subtracting alike terms depending on the operators. Figure 2 shows an example of this. The brackets in part (b) show the categorization of alike terms, while part (c) shows the result when alike terms were added together. An analysis of the high school mathematics curriculum in the Caribbean has revealed several general objectives which were found to be common among most mathematics topics. These objectives are given for all topics of the mathematics curriculum.

(a) $2y + 3z - x + 7z + y - 7x =$
(b) $(2y + y)\ (+ 3z + 7z)\ (- x - 7x) =$
(c) $3y + 10z - 8x$.

Figure 2. Simplification of Algebraic Terms

Step 3 requires the designer to select one part of the solution identified in Step 2 and match it with one of the mathematics objectives given. In the process of our analysis of the curriculum, we have identified an objective for each subtopic. Therefore, a curriculum document with each topic and its matching objective is given to the designer for designing the game.

Step 4 requires the designer to map the objective to a game type or style. At this point, the designer maps the mathematics content to the most suitable game type or style, with respect to the game play. This can be done in several ways. The designer can map the objective to their own game idea, to a game genre or a game pattern (Björk & Holopainen, 2004). In this research, we sought a technique for identifying the common successful game types which can be reused for mathematics games. A game pattern is a method of describing recurrent design choices, for example, aim and shoot, power ups, immersion, exploration and identification (Björk & Holopainen, 2004). After the analysis of the game patterns we mapped each objective to several of the most suitable game patterns.

Step 5 urges the designer to explore the chosen game style and attempt to design a game around the mathematics subtopic. Steps 6 and 7 address the two possible outcomes. If the designer has successfully derived an idea for the game then he or she is to proceed with developing the rest of the game. This is Step 8 and it recommends using a game component framework developed by Björk and Holopainen (2004) as an optional guide. If however, the designer is not able to come up with a game idea, then the designer has to find another suitable game pattern and determine if it better suits the mathematics subtopic which requires him or her to return to Step 5. In the description of each game pattern, several other related or similar game patterns are given and therefore this provides many alternative routes for the designer in deriving a game idea.

4. DISCUSSION

This paper presented an overview of the method for designing learning games for mathematics. A study will be conducted to evaluate this method for designing a mathematics learning game in an effort to gather feedback. The designer would be given the method with a detailed explanation, the mathematics curriculum which maps each subtopic to a mathematics objective, the game patterns explained, a list of game styles and the game component framework (Björk & Holopainen, 2004).

The curriculum gives a decomposed version of all the mathematics topics and each topic is mapped to a suggested mathematics objective. The game patterns are given in the form of the book by Björk and Holopainen (2004) which gives an exhaustive list of the game patterns and the explanations of each. The game designers would be asked to use the method to attempt designing a mathematics game for learning. They would be given a pre-questionnaire to get an idea of their game design experience before attempting this one. The designers would then be allowed the freedom to design a mathematics game based on high school mathematics with the method provided. At the end, they would be asked to submit any designs derived and answer the post-questionnaire which will provide feedback on the method.

In this paper, the authors have presented an efficient method for designing mathematics learning games. This method is a guide for designing a mathematics learning game in an effort to make the process easier and ensure that certain criteria are met to increase the probability for success. This method is considered improved from the traditional method because it offers a guided step by step process on how to design a game which involves the mathematics topic as a major part of the game play, as well as it is still recognized as a game by its players and not just a learning activity. It is recognized as a game as it encourages the use of popular successful game styles. The method guides the designer through the process of simplifying the mathematics topic in an effort to map it to a suitable game style. The method offers the designer several alternatives in the event that one route does not result in a game design. It is a reuse process which attempts to map a mathematics skill to a reusable successful game design style. Hence, it captures the advantages of reuse such as decreased design time, increased chances of success as a good game and less difficult to redesign new ideas each time. This method enables the designer to take advantage of aspects of other successful games and integrate mathematics into it, making learning mathematics fun.

5. FUTURE RESEARCH AND CONCLUSION

In this paper, the authors have addressed the problem of difficult game design for learning mathematics. The research involved finding a categorization for games, exploration of the mathematics topics, finding a method for mapping games to mathematics and using this to inspire the design of the game. This research has led to a unique method for designing learning games which ensures that mathematics is well-integrated into the game and increases the probability that it is a good game by reusing game styles from many other successful games. The next step of this research is to get several designers to use the method and get feedback and perhaps a finished design from this process. This exercise will also provide feedback for the improvement of the method. This method is seen as a unique contribution to game design for mathematics; however, more research is needed in order to refine and improve it with feedback from game designers.

REFERENCES

Abdullah, N. A. et al, 2008. A Toolkit Design Framework for Authoring Multimedia Game-Oriented Educational Content. *Proceedings of Advanced Learning Technologies, ICALT'08. Eighth IEEE International Conference*. 1-5 July 2008, Santander, Spain, pp.144-145.

Björk, S., and Holopainen, J., 2004. *Patterns in Game Designs*, Charles River Media Game Development: Ma, USA.

Caribexams, 2008. Caribbean Examination Council CSEC mathematics examination pass rates. Retrieved May 9, 2014, from http://www.caribexams.org/m_pass_rates

Connolly, T. M. et al, 2007. An application of games-based learning within software engineering. *British Journal of Educational Technology,* Vol. 38, No. 3, pp. 416-428.

Fanning, E., 2006. Game Mods: Customizable Learning in a K16 setting. *International Journal of Information and Communication Technology Education (IJICTE),* Vol. 2, No. 4, pp. 15-23.

Garris, R. et al, 2002. Games, motivation, and learning: A research and practice model. *Simulation and Gaming*, Vol. 33, No. 4, pp. 441-467.

Gee, J. P., 2003. What video games have to teach us about learning and literacy? *ACM Computers in Entertainment (CIE),* Vol. 1, No. 1, pp. 20-20.

Gómez-Martín, M. A. et al, 2007. Data Driven Software Architecture for Game-Based Learning Systems, In Taisch, M. & Cassina, J. (Eds.) *Proceedings of Learning with Games 2007.* Sophia Antipolis, France, pp. 229 – 236.

Gros, B., 2007. Digital games in education: The design of games-based learning environments. *Journal of Research on Technology in Education,* Vol. 40, No. 1, pp. 23-38.

Jonker, V. and Wijers, M., 2008. Th!nklets for mathematics education: re-using computer games characteristics in educational software. *Proceedings of the 8th International Conference for the Learning Sciences* ICLS'08. Utrecht, The Netherlands, 23-28 June, 2008, Vol. 1, pp. 406–413.

Ke, F., and Grabowski, B., 2007. Game playing for maths learning: cooperative or not? *British Journal of Educational Technology,* Vol. 38, No. 2, pp. 249-259.

Koster, R., 2013. *Theory of fun for game design*. O'Reilly Media, Inc., Newton, MA, USA.

Lin Y., H., 2007. Integrating Scenarios of Video Games into Classroom Instruction, *Proceedings of First IEEE International Symposium on Information Technologies and Applications in Education (ISITAE'07),* Kunming, Yuan, 23-25 November 2007, pp. 593–596.

Prensky, M., 2001. *Digital Game Based Learning.* New York: McGraw-Hill.

Rogers, S., 2010.*Level Up!: The Guide to Great Video Game Design.* NJ, USA: John Wiley & Sons.

Torrente, J. et al, 2008. Instructor-oriented authoring tools for educational videogames. *Proceedings of Advanced Learning Technologies, ICALT'08. Eighth IEEE International Conference,* 1-5 July 2008, Santander, Spain, pp. 516-518.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

PLANTS VS ZOMBIES AS AN EMPOWERING LEARNING MACHINE

Gunver Majgaard
University of Southern Denmark , Campusvej55, 5230 Odense M, Denmark

ABSTRACT

What makes the computer game Plants vs. Zombies engaging and a successful learning machine? In the game, vegetables and flowers defeat zombies by shooting peas, coconuts, and watermelons, and delaying the zombies by feeding them walnuts. The game is a classic strategic tower defence game in a vegetarian game universe. The game provides a unique example of seamlessly blending the tutorial into the gameplay.

This article analyses the game in order to identify engaging elements that can be used in the design of serious games. Empowerment and well-ordered problems were key elements found in the analysis of the gameplay in Plants vs. Zombies. Empowerment can be split into *active participation, meaningful feedback, co-design, identity, and identification.* The second key element, i.e. well-ordered problems, can be divided into *pleasantly frustrating, cycles of expertise, strategies to achieve goals, and stripped-down version.* Meaningful and engaging gameplay makes the player invests a lot of time in playing, and can trigger deep learning.

A first-hand experience of playing computer games gives valuable input for designing serious games. The first-hand experiences are also valuable in the classroom when relating game design theory to real-life experiences.

KEYWORDS

Serious games, game design, learning, computer games, empowerment, and teaching.

1. INTRODUCTION

What makes a game or serious game engaging and a successful learning machine? – How are new challenging elements introduced without reading tutorials? And how can the player exercise new evolving skills without getting bored or frustrated? The computer game Plants vs. Zombies provides a unique example of seamlessly blending the tutorial into the gameplay. Blending tutorials into the user activities can also be used in the design of serious games. The article is based on a first-hand experience of playing the Plants vs. Zombies. The target group for this article is designers of serious games, game designers and design students.

Good computer games are engaging and motivating (Gee, 2003; Fullerton, 2008; Juul, 2005; Salen & Zimmerman 2004; Kafai, 2006; Kilii, 2005; Majgaard, 2014). Matching challenges to skills is fundamental in engaging games and optimal experiences (Kilii, 2005; Csikszentmihalyi, 2005).

In serious games such as Number Blocks, Fraction Battle, and Trigonometry Apps embedding the tutorials haven’t been articulated in the design processes (Majgaard, 2009; Majgaard, 2012; Petersen, 2012). In these design processes the development teams have been focusing on how the interplay between the academic subject matter and the gameplay could improve the engaging learning processes (Habgood, 2007). And these serious games haven’t been as engaging as for example Plants vs. Zombies, and the question is, therefore whether there are specific engaging elements in Plants vs. Zombies which can be used in future serious games or used as design elements in regular games.

Gee (2005) wrote an inspiring article on games as learning machines, in which he highlighted empowered users, and understanding and problem solving as key areas for engaging the learner. In this article Plants vs. Zombies will be discussed in relation to Gee’s principles and Schön’s (1983) ideas on learning.

The research question is: What makes Plants vs. Zombies engaging and a successful learning machine - and how can this be used in designing serious games and teaching?

First, the game Plants vs. Zombies is introduced. This will be followed by a discussion of the game in relation to Gee's and Schön's ideas. Finally, the most engaging design elements in Plant vs. Zombies will be summarised.

2. PLANTS VS. ZOMBIES

The storyline in Plants vs. Zombies is thoughtful and encouraging. Zombies in computer games are usually fought with heavy weapons. In this game zombies are fought with vegetables. The game takes place in a kind of vegetarian universe. Zombies are all the time trying to destroy the ecosystem of suns and plants created by the player. The game makes the player think that planting vegetables and flowers can defeat the evil in the world (at least in the world of games). The storyline appeals to a broader target group than traditional strategy tower defence games. The game was created by PopCap Games in 2009 and sold more than 300,000 copies over the first nine days (Wikipedia Plants vs. Zombies). A sequel, called *Plants vs. Zombies 2: It's About Time*, was released in 2013. According to PopCap's webpage, 65% of their customers are female and 29% of the customers are over 29 (PopCap, 2014). In addition, PopCap Games has created games such as Bejeweled, Peggle and Solitaire Blitz.

Players place different types of plants, e.g. sunflowers (producing sunmoney), peas (shooters) and walnuts (wall), in the garden. Each one has their own unique offensive or defensive capabilities. The playing field in front of the house is divided into five horizontal lanes, and the zombies move towards the player's house along one lane. The economy in game is based on sunmoney, which can be gathered for free, and by planting e.g. sunflowers. The player starts with a limited number of seed types and at each level a new plant is introduced. The zombies also come in a number of types. They have different attributes, in particular, speed, damage tolerance, and abilities.

In the last part of the game an optional Zen garden for relaxation and contemplation is introduced. In the sequel a kitchen garden for growing extra powerful plants is introduced.

The creator George Fan intended on balancing the game between cute plants and zombies (Wikipedia Plants vs. Zombies, 2014). Strong strategic elements were included to appeal to more experienced gamers, while keeping it simple to appeal to casual gamers. Simplicity and accessibility were obtained by blending tutorials into the game (Fan, 2012). New zombies were introduced at every other level. At the first level the new zombie was presented in a safe environment where it was relatively easy to shoot, and at the next level the new type of zombie was used in a more demanding setting. This is an example of mixing tutorial and gameplay. Fan (2012) describes the peashooter as a tower with personality who nobody would expect to move. The affordance of the peashooter is intuitively understandable and defines what actions are possible (Norman, 2013).

3. DESIGN FOR LEARNING

How is learning designed into Plants vs. Zombies? Fan outlined 10 tips for making game tutorials more effective and more fun (Fan, 2012; Curtis, 2012), see the table below. In the game there is no tutorial page, the tutorial is seamlessly blended into the game. The player doesn't read about specific actions. The player is lured into doing a specific action just once. "Once they see the results of their action, that's often all it takes for them to understand that action," Fan said (Curtis, 2012). Furthermore, he minimises the use of text messages. The messaging is adaptive and if the player understands the mechanics no messages or arrows are displayed. Fan prefers to use visuals to teach, e.g. an arrow, or just showing how the peashooter works. Fan believes information creates noise, e.g. messaging on things the player already knows. Additionally, he relies on what the player already knows, e.g. that plants don't move.

Table 1. Fan's ten tips for making tutorials more effective and fun.

1. Blend the tutorial into the game	6. Use unobtrusive messaging if possible
2. Better to have the player *do* than *read*	7. Use adaptive messaging
3. Spread out the teaching of game mechanics	8. Don't create noise
4. Just get the player to do it once	9. Use visuals to teach
5. Use fewer words (max. 8)	10. Leverage what people already know

3.1 Active Participation and Empowerment

Fan's view on learning processes is based on the learner being active, e.g. "Just get the player to do it once" or "better to have the player *do* than *read*". The learning philosophy can be related to Donald Schön's ideas on learning being anchored to *active participation* and reflecting-in-action (Schön, 1983). According to Schön, the learning improves when the learner is being active and "doing" things. While the active interaction is going on, the learner reflects on the activity and how to reach the goal effectively. The player learns while interacting and the game is making *meaningful feedback*. For example, when the player for the first time plants a watermelon, he experiences how quickly a simple zombie is being defeated.

Meaningful interactivity makes the learning process more motivating and exciting. The player has influence on the gameplay and experiences the consequences of his own choices (*empowerment*).

When the player plants sunflowers, peashooters, and walnuts in front of his house, he immediately sees the consequences in terms of how successful the zombies are being kept down. While the player is observing the battle, he actively collects sunmoney for the purchase of new peashooters that immediately can be inserted in the game.

According to Gee, empowerment of learners are done by making the player a *co-designer* of the game, e.g. by planting vegetables and flowers. The player designs the garden and by that the strategy for defeating the zombies. The player chooses between a variety of plants with different abilities and characteristics, such as Jalapeno, Torchwood, Snow Pea, Power Lily, and Squash. Successful and empowered learning requires that learners feel like active agents not just passive recipients (Gee, 2003). Deep learning requires an extended commitment and such a commitment is related to *identity or identification*. In Plants vs. Zombies the player can relate to the sustainable, vegetarian, strategic and cute approach to fighting zombies.

3.2 Design of Well-Ordered Problems

According to Gee (2003), problems in good games are *well-ordered*, they shouldn't be too complex and confusing. As mentioned earlier new zombies were introduced at every other level; at the first level the new zombies were presented in a safe environment where they were relatively easy to shoot, and at the next level the new type of zombie was used in a more demanding setting. This design leads the player to form good guesses about how to proceed when he faces tougher problems later on in the game. In this sense, earlier parts of a good game are always referring to later parts (Gee, 2003).

Learning works best when new challenges are *pleasantly frustrating*. The new challenges should feel hard but doable. And even if the player fails, he should feel that he is making progress. Csikszentmihalyi (1990) describes this pleasantly frustrating relation between challenges and skills as flow. It is characterized by play and self-forgetfulness (Csikszentmihalyi, 1990). Good games adjust challenges to skills. The feedback is given in such a way that the player feels challenged and his effort is paying off. In Plants vs. Zombies the challenges progress so slowly that the player's skills adjust to the growing challenges most of the time.

Expertise is formed in any area by repeated cycles of learners practicing skills until they are nearly done automatically and can be named *cycles of expertise* (Gee, 2003). In Plants vs. Zombies each level can be described as a cycle and at each level the player can optimise his skills in collecting sunmoney, exploring new plants, and killing zombies. Good games create and support the cycle of expertise. Cycles of expertise are, in fact, the core of the gameplay in Plants vs. Zombies. Players don't like practicing skills out of context, over and over again. In good games, players learn by implementing *strategies to achieve goals* and practicing the new skills. Fan's (2012) first and second tips support this by putting focus on blending tutorials into the gameplay and learning by doing in a realistic game context. In Plants vs. Zombies the player implements strategies to achieve new goals at each level of the game while practicing new and old skills. The first levels of Plants vs. Zombies are a *stripped-down version* of the game. Only peashooters, sunflowers and walnuts are introduced. New plants and zombies are gradually introduced.

4. CONCLUSION

What makes Plants vs. Zombies engaging and a successful learning machine, and how can this be used in designing serious games and teaching? The game universe in Plants vs. Zombies makes the player think that planting vegetables and flowers can defeat the evil in the (game) world. This gives an overall meaning when playing the game. The affordance of the zombies is intuitively understandable and defines what actions are possible, e.g. a zombie with no helmet is easier to defeat than a zombie with a plastic helmet, and a zombie wearing a metal bucket is even harder to defeat. Coconut cannons hit harder than peashooters and so on.

The tutorial is seamlessly blended into the gameplay and this works remarkably well. The player is lured into doing a specific new action just once without reading any tutorial. The player might even not be aware of the seamlessly embedded learning objects. This idea has great potential in future productions.

Empowerment and well-ordered problems were key elements for deep learning and engagement of the player. They can be easily used in the designs of serious games. Design elements for empowerment are divided into the learner's *active participation, meaningful feedback, co-design, identity, and identification*. The design elements for well-ordered problems are divided into *pleasantly frustrating, cycles of expertise, strategies to achieve goals, and stripped-down versions*.

A first-hand experience of playing computer games gives valuable input for designing serious games. The first-hand experiences are valuable in the classroom when relating theory to real-life experiences.

REFERENCES

Csikszentmihalyi, M., (1990). Flow: The Psychology of Optimal Experience. New York: Harper and Row. ISBN 0-06-092043-2

Fan G., 2012. How I Got My Mom to Play Through Plants vs. Zombies. GDC Game Design. http://www.gdcvault.com/play/1015541/How-I-Got-My-Mom retrieved 2014-30-12

Fullerton, T., (2008). Game Design Workshop. A playcentric approach to creating innovative games. Morgan Kaufmann

Gee, J. P., 2003. What Video games have to teach us about learning and literacy. New York: PalGrave-McMillan.

Habgood, M. P. J., (2007): The effective integration of digital games and learning content. Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy July 2007

Juul, J. (2005). Half-real: video games between real rules and fictional worlds. Cambridge, MA: MIT Press

Kafai, Y. B., (2006). "Playing and Making Games for Learning Instructionist and Constructionist Perspectives for GameStudies". Games and Culture Volume 1 Number 1. January 2006 36-40. Sage Publications.

Kiili K. 2005,Digital game-based learning: Towards an experiential gaming model. The Internet and Higher Education Volume 8, Issue 1, 1st Quarter 2005, Pages 13–24, Elsevier, doi:10.1016/j.iheduc.2004.12.001

Majgaard, G. 2014. The playful and reflective game designer. *Electronic Journal of E-Learning*. 12, 3, s. 271-280

Majgaard, G, Nielsen, J & Misfeldt, M., 2012, 'The Learning Potentials of Number Blocks'. *Towards Learning and Instruction in Web 3.0. Advances in Cognitive and Educational Psychology*. Springer, s. 289-302.

Majgaard, G 2009, 'The Playground in the Classroom - Fractions and Robot Technology'. . *Proceedings of Conference of CELDA* Cognition and Exploratory Learning in Digital Age. International Association for Development, IADIS, s. 10-17.

Pedersen, JB, Andersen, S & Majgaard, G., 2012, 'Design of Trigonometry Apps for Vocational Education'. *Proceedings of IASTED International Conference of Human Computer Interaction HCI.*, 10.2316/P.2012.772-020

PopCap, 2014. Mediaroom. http://popcap.mediaroom.com/index.php?s=30842 retrieved 2014-12-30

Salen, K., Zimmerman E., (2004). Rules of Play: Game Design Fundamentals. MIT Press 2004

Sicart, M., (2008). "Defining Game Mechanics." The International Journal of Computer Game Research, volume 8 issue 2 December 2008, ISSN:1604-7982

Wikipedia Plants vs. Zombies. 2014. http://en.wikipedia.org/wiki/Plants_vs._Zombies retrieved 2014-12-17

Interview with Plants vs. Zombies creator George Fan. http://www.gamezebo.com/2009/05/22/interview-plants-vs-zombies-creator-george-fan/

Norman D. A., (2013). The Design of Everyday Things Paperback. Basic Books

Schön D., 1983. The Reflective Practitioner: How professionals think in action. London: Temple Smith, 1983.

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

EXTRACTING PUPIL CONTOUR WITH HOUGH TRANSFORM OF CONNECTED COMPONENT BOUNDARIES

Ivan A. Matveev[1], Nikolay N. Chinaev[2] and Vladimir P. Novik[3]
[1]*Computing Center of Russian Academy of Sciences, Vavilov str., 40 Moscow, 119333, Russia*
[2]*Moscow Institute of Physics and Technology, Institutsky per., 9, Dolgoprudny, Moscow region, 141700, Russia*
[3]*Iritech Inc, Moscow R&D office, B.Tatarskaya str., 21, Moscow, 115184, Russia*

ABSTRACT

A method for determining the pupil boundary in the image of eye is proposed. The method is based on examining connected components of binarized image. The components are first sorted according to probability to contain pupil, which is estimated from their inertia moments by empirically derived rules. Hough transform is applied to border of the component and outlines the center of most pronounced circular arc. If center hypothesis is strong enough the histogram of distances of border points to this center is built. If pronounced maximum exists on the histogram it means that a circle or an arc is present in this component, which is taken for pupil boundary. The method benefits from both stability of Hough approach and small number of pixels detected as boundaries of connected components. Tests were performed with several iris image databases from public domain.

KEYWORDS

Iris Segmentation, Hough Transform.

1. INTRODUCTION

Iris region segmentation in eye images is an interesting and actual task for biometric identification and medical applications. An important sub-task is detection of pupil boundaries that is in fact search for rounded dark object. Circle detection is one of the most popular problems in image processing field. Numerous different methods have been developed, some of which are applicable in the particular case of pupil border detection. One can outline morphology methods, described by Bowyer et al (2008), projections of brightness and its gradient (Matveev, 2010 and Mohammed et al, 2010), optimal contour construction (Ritter and Cooper, 2003). Big share of methods (Chen et al. 2012, Ma et al 2002, Matveev 2012, Pan et al. 2011) relies on Hough transform ideology, proposed back in 1959 by Paul Hough. The advantage of Hough transform is its stability against partial occlusion of the figure, for example, it can be used to find the circumference by one of its arcs. The drawback is a relatively high computational cost arising from complex processing of candidates. Object detection methods based on morphology and brightness projections often apply brightness thresholding (or more elaborate techniques) as the first step to create binarized image, which is then processed. This is done under assumption that pixels of object region have uniform characteristics, which are distinct from those of background. Since binarization and extracting connected objects are computationally simple procedures, the advantages of this method is the execution speed. However, quite often binarization generates several connected components, as the brightness of the eyebrows or eyelashes may have same brightness as the pupil. Thus eyelashes and eyelids may merge with the pupil to form one component of inadequate form. Therefore, the disadvantage of this class of methods is a significant share of errors associated with the distortion of the shape of object.

The presented work attempts to combine the advantages of the method of binarization, which retrieves a small number of points of interest in the image, and Hough transform, which extracts the shape of a given type from very noisy data. On the binarized image connectivity components are outlined, boundary of each is subjected to Hough transform to find the rounded contours. This combination of methods is not known to the authors from literature.

2. PROBLEM STATEMENT AND SOLUTION METHOD

Image of eye with the surrounding parts of the face (eyebrows, nose, cheeks part) is considered. It is necessary to find a circle, which is the best approximation of the boundary between pupil and iris. The border can be partially occluded by eyelashes, eyelids, reflections. Pupil is dark but not necessarily darkest object in image. The workflow of method is presented in the block diagram, see Fig.1.

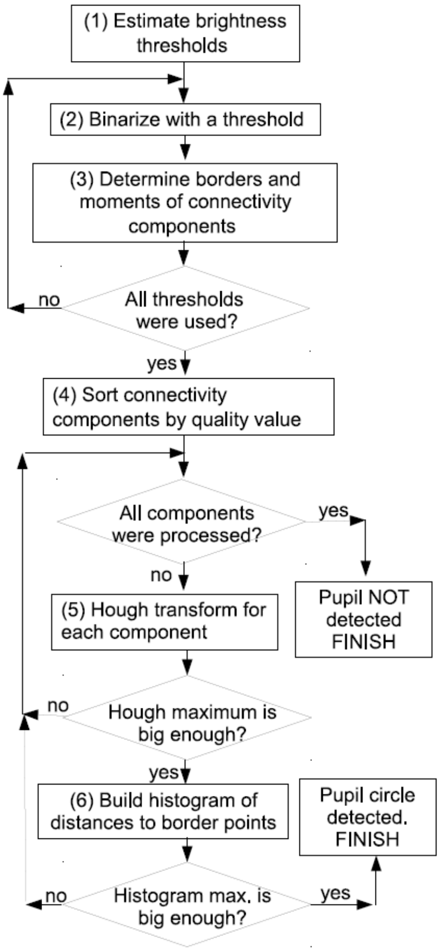


Figure 1. Block diagram of method functioning

Now we describe it briefly and in the following subsections more details are given. Several thresholds of binarization are calculated from image brightness histogram (block 1). Connectivity components corresponding to the dark areas are outlined by binarization (block 2). For each component its boundary and inertia moments are determined (block 3). Moments of the component are used to calculate the likelihood that pupil belongs to a this component, then components are sorted so as to process ones with highest likelihood first (block 4). The boundary of each component is a cyclic sequence of pixels. It should be noted that the number of border pixels is substantially smaller than their total number in the image or the number of pixels with high brightness gradient, which could be considered as parts of boundary. The pupil not always has border with brighter iris only, but is often covered with eyelids or eyelashes that causes distortions of two kinds: the boundary of the pupil is visible partially; pupil merges with eyelids or eyelashes. In this case only a small part of detected boundary belongs to pupil, however it is necessary to determine visible part of the boundary. Hough transform is performed (block 5), which includes construction of a beam in the accumulator for each pixel of boundary. As a result, if the boundary contains circular arc, significant local maximum arises in the accumulator at the center of hypothetical circle containing this arc. Global maximum of accumulator corresponds to the most likely position of the circle. To determine the radius of the circle histogram of distances from the found center to the boundary pixels is built (block 6). Position of the histogram maximum gives the radius. Following subsections describe steps of the method in more detail.

2.1 Binarization and Characterizing Connectivity Components (Blocks 1,2,3)

As the pupil is dark image area, it can be outlined as a connectivity component when the binarization threshold value exceeds pupil brightness and is below iris brightness. Boundary of this connectivity component is a pupil border. As levels of brightness of pupil and iris are unknown a-priory, and hence the threshold dividing them is also unknown, several passes (denoted their count as K) of algorithm with different thresholds are performed. The values of applied thresholds are determined so that the number of pixels with brightness below the threshold is a certain fraction of the total pixel number. For this purpose brightness histogram is used:

$$h(b) = \left|\{\vec{p} : I(\vec{p}) \le b\}\right| \; . \qquad (1)$$

Pupil is a dark object, so histogram part corresponding to the low brightness is analyzed. Shares of points α_k, determining the binarization thresholds, are set uniformly in range from 0 to 1/4 of total pixel count:

$$\alpha_k = \frac{1}{4}\frac{k}{K} \; , \; k = 1,...,K \; . \qquad (2)$$

Binarization thresholds B_k are calculated by histogram (1) and shares (2):

$$B_k : \; h(B_k) = \alpha_k WH \; , \qquad (3)$$

where W and H are width and height of the image in pixels.

Binarization with threshold B is performed to zero brightnesses over the threshold and outline the pixel with lower intensities.

In the binarized image dark and bright areas form several connectivity components. We consider the components of dark pixels, and treat bright pixels as background. Here 8-connectivity is used i.e. pixel $\vec{q}$, $I_B(\vec{q}) = 1$ is considered connected with pixel $\vec{p}$, $I_B(\vec{p}) = 1$, if $\vec{q}: \; |p_x - q_x| \le 1, \; |p_y - q_y| \le 1$. For each connectivity component C ordered list of border points $L(C) = \{\vec{l}_i\}$ and moments of zero order $M(C)$, first order, and second order are calculated.

2.2 Rejection and Sorting of Connectivity Components (Block 4)

Components detected by binarization are subjected to Hough transform. However, some of them can be discarded before the processing merely by the statistical characteristics as obviously not containing the pupil, and the remaining can be sorted, so that the first processed components contain pupil with the highest probability. To remove components, obviously not containing the pupil parameters of *equivalent ellipse* are considered, where a, b, φ are big and small half-axes and direction of big half-axis respectively. Connectivity component containing pupil should have a sufficiently large area, and the ratio of the major and minor semi-axes must lie within a certain range. Furthermore, if eyelashes/eyelids have same or smaller brightness than the pupil, induced distortion of the connected component extends it horizontally but not vertically. Accordingly, the objects with equivalent ellipse having large eccentricity and major axis located closer to the vertical should be discarded. These conditions are written as:

$$ab > \rho_{\min}^2 \; ; \; \frac{a}{b} < T_1 \; ; \; \neg\left(\left(\frac{a}{b} > T_2\right) \cap \left(\varphi \in \left[\frac{\pi}{4}; \frac{3\pi}{4}\right]\right)\right), \qquad (4)$$

where $\rho_{\min}$ is minimal possible radius of pupil in pixels. Optimal values of thresholds $T_1 = 5$, $T_2 = 2.5$ are determined experimentally.

To sort residual components the following quality criteria is calculated:

ISBN: 978-989-8533-38-8 © 2015

$$\Theta(C) = \frac{M(C) - l(C)}{l^2(C)} \quad , \tag{5}$$

where $l(C) = |L(C)|$ is a length of connected component border. This feature increases as the shape of the region becomes more circular and also increases with the size of the area. Thus, connected components are sorted by descending characteristic (5), which favors rounded components of significant size.

2.3 Hough Transform (Block 5)

Voting procedure is performed as follows: inner normal $\vec{\eta}_i$ to each pixel $\vec{l}_i \in L$ are constructed, a segment of normal enclosed in range $[\rho_{\min}; \rho_{\max}]$ is selected, counter function in appropriate accumulator points is increased.

$$A(\vec{u}) := A(\vec{u}) + v(\vec{p}, \vec{l}_i),$$
$$v(\vec{u}, \vec{l}_i) = \begin{cases} 1, \, if \; \vec{u} = r\vec{\eta}_i + \vec{l}_i \, , \; r \in [\rho_{\min}; \rho_{\max}], \\ \qquad 0, \, otherwise. \end{cases} \tag{6}$$

Constructing normal vector $\vec{\eta}$ at the pixel $\vec{l}$ is done as follows: closest neighbors of pixel in list L are selected, define them as $\{\vec{l}_j\}$, where negative index corresponds to moving backward in the list and positive is forward. Tangent direction $\vec{\tau}$ is estimated from the neighbor pixel coordinates:

$$\vec{\tau} = \sum_{j=-n}^{-1} \frac{\vec{l}_i - \vec{l}_j}{|\vec{l}_i - \vec{l}_j|} + \sum_{j=1}^{n} \frac{\vec{l}_j - \vec{l}_i}{|\vec{l}_j - \vec{l}_i|} \; , \tag{7}$$

where n is a size of pixel neighborhood and is equal to $\rho_{\min} / 2$. From two possible normal directions $\vec{\eta}^{(1)} = (\eta_x; \eta_y) = (-\tau_y; \tau_x)$ and $\vec{\eta}^{(2)} = (\tau_y; -\tau_x)$ inner normal is selected according to direction of border enumeration. Segment for voting is constructed in rastered accumulator by Bresenham algorithm. After the voting the accumulator is smoothed by the low-pass filter G and center of the hypothetical circle is determined as the global maximum in smoothed accumulator:

$$\vec{c} = \arg\max_{\vec{u}} \left(A(\vec{u}) * G \right) \, . \tag{8}$$

2.4 Building Histogram of Distances to Border Pixels (Block 6)

A histogram $h(r)$ of distances from center $\vec{c}$ and all pixels of the border $L = \{\vec{l}_i\}$ is constructed:

$$h(r) = \left| \{ \vec{l}_i : \vec{l}_i \in L, \; r - 0.5 \le |\vec{c} - \vec{l}_i| < r + 0.5 \} \right| \, . \tag{9}$$

Histogram has a specific properties in case when pupil belongs to the component and its center is correctly detected. A peak position in the histogram corresponds to the radius of pupil in image: $r_P = \arg\max_r h(r)$. The final conclusion about pupil presence is based on analysis of $h(r)$. Mass of the histogram is calculated in window of width w centered at r_P, which is position of histogram maximum. The resulting value is an estimate of the length of the visible contour of the pupil, and is compared with r_P. Is is considered that pupil is detected if

$$\sum_{r=r_P-w}^{r_P+w} h(r) > \pi r_P \; , \tag{10}$$

i.e. sufficient condition for pupil detection is visibility of half of its border. The width w of the summing window depends on precision of center detection, it is set here to $w = 2$ according to numerical tests.

2.5 Estimation of Algorithmic Complexity

Image size if $W \times H$, $W \sim H$, and the width of image W can be considered as image linear size, thus number of image pixels is proportional to W^2.

Estimation of binarization thresholds includes building and processing brightness histogram with complexity $\max\{W^2; L\}$, where L is a number of brightness levels in image. Grayscale 8-bit images are used hence $L = 256 << W^2$. Binarization and connectivity component selection for one binarization threshold have complexity W^2. Number of connectivity components is limited by $W^2 / \rho_{\min}^2$, where $\rho_{\min}$ is a minimum expected pupil radius. Length of the border can be evaluated as W. Total complexity of K passes of binarization is KW^2, as a result $KW^2 / \rho_{\min}^2$ connectivity components are extracted. Hough voting for one pixel has complexity of W, and for whole border it is W^2. Maxima search in accumulator requires W^2 operations. Thus, considering the above total complexity of the method comprises $KW^4 / \rho_{\min}^2$.

With the increase of image resolution and its linear size W expected pupil size also grows linearly, so one can state that $\rho_{\min} \sim W$ and the complexity of the algorithm is in fact KW^2. It grows linearly with respect to image pixel number. Direct application of Hough transform to locate circle yields complexity of W^3. The reduction of complexity in the proposed approach is produced by using preselected border pixels, which number is proportional to W, instead of treating all image points (or some share of them selected by thresholding), which number is proportional to W^2.

3. EXPERIMENTS

Four public domain iris image databases were used for tests, defined as: BATH (Monro, 2005), CASIA (Chinese academy of Sciences, 2005), MMU (Multimedia university, 2006), NDIRIS (Phillips et al., 2010). These images were marked by a human expert: in each a circle approximating pupil border was set. Define center of such circle as $(x_0; y_0)$ and its radius r_0. Algorithm detected center (x, y) and radius r of pupil circle in each image and these data were matched against expert ones, the deviation of center was calculated as $Q_C = |x_0 - x| + |y_0 - y|$ and the deviation of radius as $Q_R = |r_0 - r|$. Table 1 represents comparison of the method with its analogues. The following characteristics are given: average deviation of pupil center QC (average value of Q_C for all test set), average radius deviation QR, both expressed in pixels, and mean calculation time T, expressed in milliseconds.

Table 1. Comparison with other methods

Database	BATH			CASIA			MMU			NDIRIS		
Values	QC	QR	T	QC	QR	T	QC	QR	T	QC	QR	T
Masek, 2003	5.32	6.72	108.2	3.67	5.15	97.52	4.98	5.78	99.78	5.59	7.23	112.2
Ma et al., 2004	4.29	4.65	376.8	4.79	5.39	363.6	3.92	4.67	317.1	5.92	7.38	378.8
Chen, 2012	3.27	3.19	26.55	1.19	3.02	29.17	1.14	3.76	25.47	1.79	3.11	27.61
Proposed	4.61	1.59	97.28	2.82	3.26	93.95	3.11	0.99	25.22	3.94	1.89	98.30

4. CONCLUSION

Method of border contour detection in eye image is proposed. A contours are built from pixels of boundaries of connectivity components obtained from dark regions in source image. To select correct component Hough transform of contour is used. The transform gives a center of circular part of component. Radius of the circle is defined as the maximum of the histogram of distances of boundary pixels to a dedicated center. Then pupil border pixels may be selected from the set of contour pixels as being close to that radial distance to the center.

Tests of the algorithm were performed on sets of images of public domain databases. Using Hough transform one can correctly locate the true center, radius and contour of the pupil, even in the presence of high interference, provided only partially visible contour of the pupil. The disadvantage of this method is potentially great execution time, which may occur when iterating through many connected components and several binarization thresholds. This drawback is partially mitigated by using quality measure of connectivity components.

ACKNOWLEDGEMENT

The work was funded by RFBR grant No.15-01-05552.

REFERENCES

Bowyer, K., Hollingsworth, K., and Flynn, P., 2008. Image understanding for iris biometrics: a survey. *In Computer Vision and Image Understanding*, 110(2):281 307.

Chen, D., Bai, J., and Qu, Z., 2012. Research on pupil center location based on improved Hough transform and edge gradient algorithm. *In Proc. National Conf. Information Technology and Computer Science*, page 47-51.

Chinese academy of sciences institute of automation, 2005. CASIA Iris image database version 3. *In http://www.cbsr.ia.ac.cn/ IrisDatabase.htm* .

Daugman, J., 2007. New Methods in Iris Recognition. *In IEEE Trans. Systems, Man and Cybernetics-Part B: Cybernetics*, V.37. P.1167-1175.

Hough, P., 1959. Machine analysis of bubble chamber pictures. *In Conf.Proc., C590914*:554–558.

Ma, L., Tan, T., Wang, Y., and Zhang, D., 2004. Local intensity variation analysis for iris recognition. *In Pattern Recognition*, 37(6):1287–1298.

Ma, L., Wang, Y., and Tan, T., 2002. Iris recognition using circular symmetric filters. *In Proc. 16th Int. Conf. Pattern Recognition*, 1(2):414 –417.

Masek, L., 2003. Recognition of human iris patterns for biometric identification. *In http:// www.csse.uwa.edu.au/ pk/ studentprojects/ libor*.

Matveev, I. A., 2010. Detection of iris in image by interrelated maxima of brightness gradient projections. *In Applied and Computational Mathematics*, 9(2):252 –257.

Matveev, I. A., 2012. Iris center location using Hough transform with two-dimensional parameter space. *In Journal of Computer and Systems Sciences International*, 51(6):785 – 791.

Mohammed, G., Hong, B., and Jarjes, A., 2010. Accurate pupil features extraction based on new projection function. *In Computing and Informatics*, 29:663 – 680.

Monro, D. University of Bath Iris Image Database. *In http:// www.bath.ac.uk/ elec-eng/ research/ sipg/ irisweb/* .

MMU Iris Image Database: Multimedia University. *In http:// pesonna.mmu.edu.my/ ccteo/* .

Pan, L., Chu, W., Saragih, J., and de la Torre, F., 2011. Fast and robust circular object detection with probabilistic pairwise voting. *In Signal Processing Letters*, 18(11):639–642.

Phillips, P., Scruggs, W., O'Toole A. et al., 2010. Frvt2006 and ice2006 large-scale experimental results. *In IEEE PAMI*, 5(32):831–846.

Ritter, N. J. and Cooper, J. R., 2003. Locating the iris: A first step to registration and identification. *In Proc. 9th IASTED Int. Conf. Signal and Image Processing*, pages 507–512.

COMPUTER-AIDED TOOL FOR BREAST TUMOUR ANALYSIS PURPOSES

Guillermo Palacios-Navarro
University of Zaragoza, Teruel, Spain
Universidad de Las Fuerzas Armadas ESPE, Sangolqui, Ecuador
Prometeo Project Researcher (SENESCYT), Ecuador

ABSTRACT

This work introduces a tool specifically designed to help pathologists in their daily work. The software tool integrates several applications related to different tumoral parameters very useful to analyze and detect tumours. All the proposed algorithms have in common the use of the most suitable colour space in every case in order to isolate the different tissues and therefore, to perform an appropriate segmentation. Results are presented for every single application, showing a great deal of accuracy in most cases, in comparison with that obtained by specialists in a manual way. The tool contributes to a higher level of assurance when estimating the seriousness of a tumour and its subsequent diagnosis.

KEYWORDS

Medical imaging, image segmentation, tumor detection, computer-aided system.

1. INTRODUCTION

Image segmentation is a fundamental part in many applications in the area of image processing, which entails the division of images into regions of similar attribute. Meaningful segmentation in medical imaging is still difficult and complex, due to the inherent noise and poor contrast.

Selecting the suitable colour space is the first step to accomplish colour segmentation. There are several well-known colour spaces for image representation and display. Although the perceptual colour spaces (CIE or HSV) may be more suitable for image representation and analysis, the selection of a colour space for image processing is image/application dependent. There is no any colour space better than others and suitable to all kinds of images yet.

A key problem in medical imaging is automatically segmenting an image into its constituent heterogeneous processes (Corso, 2008). Pathologists usually make diagnostic decisions by viewing specimen cells and assessing their various diagnostically important parameters such as size, shapes and textures (Gil, 1994). Accurate measurements of the cell parameters with a computer aided image processing system are also useful in quantitative pathology and medicine. To obtain these critical parameters accurate segmentation of the nuclear regions is required. Breast cancer is the most frequently diagnosed cancer and a leading cause of cancer deaths in women. The development of computer-aided detection (CAD) systems has shown its efficacy to improve diagnostic accuracy in breast cancer and therapeutic planning.

Literature is plenty of CAD systems. For instance, we can find examples in the field of ultrasonography, which is widely used for the detection and evaluation of many diseases, such us breast tumours. Huang integrated the advantages of neural network classification and morphological watershed segmentation to extract precise contours of breast tumours from ultrasound images (Huang, 2004). Alvarenga et al. (Alvarenga, 2007) classified breast tumors on ultrasound images, using texture features calculated from complexity curve and grey-level co-occurrence matrix applied to a hybrid classifier based on a multilayer perceptron network and genetic algorithms.

Computer-Aided systems have the potential to positively impact clinical medicine by freeing specialists from the burden of manual labelling and by providing more robust measures to aid in diagnosis of diseases. Pathologists find it very interesting and useful for their diagnosis automatic calculations once image has been segmented.

The main contribution of this paper deals with the creation of a simple by robust software tool to facilitate the daily work to pathologists. It has been designed aimed at facilitating in an automatic or semi-automatic way, valuable information about tumour detection and/or progression. This paper is organized as follows. Section 2 and 3 introduce the different developed applications, together with representative results. Finally, section 4 presents the main conclusions and the future work.

2. CELL NUCLEI SEGMENTATION

2.1 Introduction

Immune-histochemistry is the location of antigens in tissue sections by using labelled antibodies as specific reagents through antigen-antibody interactions that are visualized by a marker. It has become a crucial technique and widely used in many medical research laboratories as well as in clinical diagnostics.

Nevertheless, one of the current problems working with these techniques deals with the interpretation of the obtained results. Interpretation errors decrease to an acceptable level when pathologists have enough experience in such techniques. Normally they work on a manual basis, by means of tedious processes, so results are individual-dependent and for this reason it is necessary to create an automatic process to increase accuracy. Figure 1 shows a breast tumour treated with an immune-histochemical technique. Cells (brown colour) are easily distinguished from the rest of tissues, such as adipose cells (white colour) and stroma (blue colour).

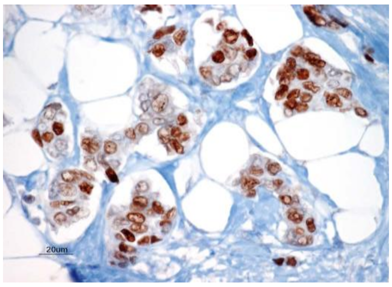


Figure 1. Original image treated with an immune-histochemical technique. Brown (dark) colouring cells have a certain quantity of antigens (estrogens).

2.2 Procedure

In order to achieve a good level of segmentation, it is necessary to bear in mind that cells exhibit brown colouring, whereas adipose cells and stroma exhibit white and blue colouring, respectively. Colour spaces HSV, LUV and YIQ will be used. Six different kinds have been defined: stroma, adipose cells, remaining parts of tissue and three levels for coloured cells. A classification based on Euclidean distance will be used to segment them and a refining stage will be necessary to improve accuracy. Figure 2 shows the general schema showing the segmentation process.

The LUV colour space will allow us to separate colour information from intensity, being efficient to control little colour differences. It can be observed that pixels having an L value below 0.4 belong to adipose cells, whereas pixels having an L value above 0.4 belong to the rest of tissues. As far as estrogen cell segmentation is concerned, we have used the Euclidean metric, mainly because of its rotation invariance property. Its effectiveness has been proved in many applications (Liao, 2006) (Wirth, 2008).

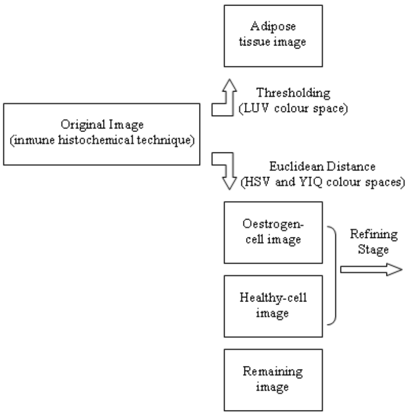


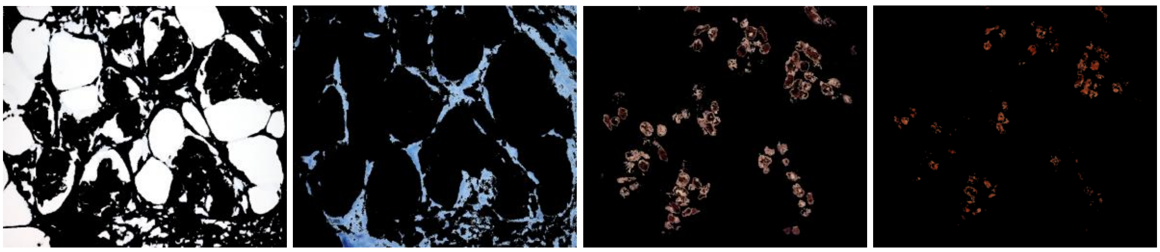
Figure 2. General schema showing the segmentation process

Beginning our analysis within the HSV colour space, it has been checked that cells maintain a uniformity in the hue component (H), differing from the rest of tissues. We also realized that saturation component (S) is not always reliable due to a hue instability caused when its value is close to zero (pixels having low saturation are left unassigned to any regions). Results using this component were not as good as expected, so we searched for a second parameter in the YIQ colour space, obtaining optimal results by considering the I component. Therefore, these two parameters will allow us to distinguish and classify cells in three different levels according to the quantity of colouring absorbed as well as to distinguish them from the rest of tissues.

The most critical parameter to achieve a good segmentation in the Euclidean distance algorithm deals with the choice of the geometrical centres (centroids) representing every single kind of tissue. The Euclidean Distance among pixels of the entire image and the HC and IC centroids of the region are obtained following the equation: $d = \sqrt{(H - H_C)^2 + (I - I_C)^2}$.

By calculating the Euclidean Distance among the pixels belonging to the original image and centroids, we obtain six different matrixes, each defining a different tissue.

In order improve the accuracy in the segmentation process, some pixel re-allocation tasks should be performed to avoid some artifacts such as non-uniformity in the saturation mean value near the boundaries of cells (refining stage). Figure 3 shows the result of the segmentation process. High level and low level colouring images should be superimposed in order to get the total quantity of colouring.



(a) (b) (c) (d)

Figure 3. Segmentation process results for figure 1. (a) Adipose cells. (b) Stroma. (c) Low-level colouring cells. (d) High-level colouring cells

2.3 Results

Once tumoral cells (cells having the estrogens antigen) have been isolated, positivity grade will be defined as the percentage of cells containing the antigen with regard to the total number of them. This figure is normally estimated by pathologists in a manual way, by means of a tedious counting process.

To estimate the positivity grade, we will divide the total number of pixels belonging to classes 2 and 3 (cells containing the estrogens) by the total number of pixels belonging to cells tissues (levels 1, 2 and 3). Table 1 shows a comparison of results obtained between the traditional method and our automatic method.

Table 1. Estrogens grade estimate (Manual vs. Automatic)

Cell images	Manual method	Automatic method	error
Img_ 1548	**25,7 %**	**29,3 %**	**3,6 %**
Img_ 1549	**40,9 %**	**45,3 %**	**4,4 %**
Img_ 1757	**35,4 %**	**39,1 %**	**3,7 %**
Img_ 1758	**40,8 %**	**44,8 %**	**4 %**
Img_ 1759	**45,8 %**	**49,3 %**	**3,5 %**
Img_ 1857	**43,1 %**	**44,1 %**	**1 %**
Img_ 1858	**46,4 %**	**49,8 %**	**3,4 %**
Img_ 1859	**25,7 %**	**28,2 %**	**2,5 %**
Img_ 2049	**39,2 %**	**41,3 %**	**2,1 %**
Img_ 2050	**32,7 %**	**34,2 %**	**1,5 %**

The proposed algorithm is a simply but robust tool, providing a high calculation speed. The results are very close to those obtained by specialists in a manual and not very friendly way. Errors made with our proposed method are below +5 %, being within the range specialists work, and always positive, which contributes to a higher level of assurance when estimating the seriousness of the tumour and its subsequent diagnosis.

3. NEURAL NETWORK-BASED TUMOR DETECTION

3.1 Introduction

The analysis of the different parameters of cells will allow us to perform a deeper analysis of tissues. The images considered have been taken directly from samples of extracted tumors. Our aim lies on quantifying variations of cells in size, shape, chromatism, etc. For this purpose, we have a wide database of photographs belonging to both healthy cells and tumoral ones. We have calculated a series of parameters in order to perfectly characterize both healthy and tumoral cells. Some of the characteristics or texture parameters that can be used to describe quantitatively regions segmented in a certain image can be divided into two categories: geometric and chromatic characteristics. The geometric characteristics contain information of shape, position, size and orientation of a certain region (such us area, perimeter and eccentricity) and will serve us to characterize adequately the cells object of study. On the other hand, we will use some chromatic characteristics for the H component in the HSV colour space, such as: mean value, mean value of gradient on the border, mean value of second derivative, local variance, contrast, correlation, entropy, energy.

Taking into account the statistical study carried out over a wide range of variables, together with the pathologist recommendation based on experience, we have chosen a neural network for classification purposes. The so chosen input variables for our NN (independent variables) result in a great deal of variability on the NN output variables (dependent variables). We consider that the number of variables (10) to be used is good enough to reach reasonable results. Thereby avoiding a high bias and a high generalization error, the input variables consist on geometric and chromatic characteristics.

3.2 Neural Network Design

Based on the approach of Funahashi (Funahashi, 1989), there is theoretical evidence for which it is enough with a hidden layer to solve any problem of functional approach with a perceptron network, though sometimes it is resorted to two by practical questions. Therefore, we have chosen only one hidden layer. To decide the number of hidden neurons for this layer, we will state initially for the trial and error method, that is to say, beginning for a small number of neurons (4, for example) and increasing the value little by little having a training with cross-validation, so as to reach the suitable architecture in terms of the optimum result for the sets of learning and test patterns. We have obtained optimum results for an architecture consisting of 7 neurons in the hidden layer. Our network will also consist of 10 input neurons (corresponding to the ten chosen parameters) and an output neuron. The output neuron will provide a value close to 1 if the selected cell is a non tumoral one (healthy) and a value close to -1 if the cell is a tumoral one.

3.3 Results

The set of learning consisted of approximately 300 cells, chosen at random among all the tissues samples. We have taken 80 % for training, and the 20 % remaining for test. The so trained NN has been used to classify more than 100 different cells belonging to diverse breast tissues (both healthy and tumoral), giving an efficiency of 100 %. Neither false positives nor false negatives were found at all.

4. CONCLUSION

In this work, we have integrated several segmentation algorithms to build a tool suitable for the analysis of tumour and guiding pathologists to a more straightforward diagnosis. The obtained results in every application showed a high level of accuracy, providing a valid tool for pathologists in the field of prevention of tumoral diseases and therapeutic planning. Future work is concerned with the development of new applications in our CAD system to evaluate the histological grade in breast tumours as well as the analysis extrapolation to other kind of tumours. The processing results presented in this work have been obtained using Matlab®.

ACKNOWLEDGEMENT

Authors gratefully acknowledge the support of Prometeo Project, Secretaría de Educación Superior, Ciencia, Tecnología e Innovación from Ecuador, as well as the Fundación Antonio Gargallo for its support under the grant 2014/B017. Authors also thank their clinical partners from the Pathological Anatomy department, at Obispo Polanco Hospital in Teruel, for their support and valuable contribution.

REFERENCES

Alvarenga, A.V. et al., 2007. Classifying Breast Tumours on Ultrasound Images Using a Hybrid Classifier and Texture Features. *Proceedings of International Symposium on Intelligent Signal Processing*, Alcalá de Henares, Spain, pp. 1-6.

Corso, J. J. et al., 2008. Efficient multilevel Brain Tumor Segmentation with integrated Bayesian Model classification. *IEEE Transactions on Medical Imaging*, vol. 27, No. 5, pp. 629-640.

Funahashi, K. I., 1989. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2, pp. 183-192.

Gil, J. and J. Barba, J., 1994. Morphometry in image analysis for anatomic pathology. In Image Analysis: A Primer for Pathologists. A. M. Marchevsky and P. H. Bartels, Eds. Raven, NY, pp. 79-124.

Huang Y.L. and Chen, D.R., 2004. Watershed segmentation for breast tumor in 2-D sonography. *Ultrasound in Medicine & Biology*, Vol. 30, No. 5, pp. 625-632.

Liao, M.Z.W. et al., 2006. An Affine Invariant Eulidean Distance Between Images, *International Conference on Machine Learning and Cybernetics*, pp. 4133 – 4137.

Palacios, G., Beltran, J.R., 2007. Cell nuclei segmentation combining multiresolution analysis, clustering methods and colour spaces. *Proceedings of the IEEE International Conference Machine Vision and Image Processing*. Kildare, Ireland, pp. 91–97.

Palacios, G., Beltran, J. R., 2008. A wavelet transform based multiresolution edge detection and classification schema. *Proceedings of SPIE 7000, Optical and Digital Image Processing, 70000U*, Strasbourg, France.

Wirth, W., and Eckstein, F., 2008. A Technique for Regional Analysis of Femorotibial Cartilage Thickness Based on quantitative Magnetic Resonance Imaging. *IEEE Transactions on medical imaging*, vol. 27, No. 6, pp. 737-744.

ISBN: 978-989-8533-38-8 © 2015

IMAGE SAMPLING AND RECONSTRUCTION USING COMPRESSIVE SENSING

Guoqing Wu, Wengu Chen and Yi Cao
Institute of Applied Physics and Computational Mathematics, No.2 Fenghao East Road, Haidian District, Beijing, China

ABSTRACT

There has been growing interest in unifying the fields of compressive sensing and sparse representations to perform imaging. In this paper, we have reviewed compressive sensing theory and studied the scheme of image sampling and reconstruction. The whole process measures a subset of the pixels in the photograph and uses compressive sensing algorithms to reconstruct the entire image from this data. We have also analyzed and compared the combination influences of various sensing and sparse transform matrices, subsampling rate and recovery algorithms. Experimental results are very encouraging and constructive, both visually and quantitatively. From the results, we have concluded that Restricted Isometry Condition (RIC) plays an important role in the quality of the reconstructive images. The results also clearly demonstrate the efficacy of the compressive sensing in image reconstruction.

KEYWORDS

Compressive Sensing, Sparse Representation, Sampling and Reconstruction.

1. INTRODUCTION

Real-world images produced by digital camera devices are compressible in transform domains, which is the reason for the success of transform-coding compression algorithms such as JPEG and JPEG2000. However, most imaging systems do not take advantage of this compressibility when capturing the image. Instead, they measure and store the information at every pixel and then throw out most of this information during the compression process. Naturally, this raises the question if we can sample only the “important” information in an image directly without wasting effort (time, power, bandwidth, etc.) measuring data that will be thrown away eventually during compression.

Image sampling and reconstruction aims to reconstruct a high quality image $\boldsymbol{x}$ from its degraded measurement y . It is a typical ill-posed inverse problem and it can be generally modeled as

$$y = DHx + \upsilon \tag{1}$$

where $\boldsymbol{x}$ is the unknown image to be estimated, H and D are degrading operators and υ is additive noise. When H and D are identities, the problem becomes denoising; when D is identity and H is a blurring operator, it becomes deblurring; when D is identity and H is a set of random projections, it becomes compressive sensing; when D is a down-sampling operator and H is a blurring operator, it becomes (single image) super-resolution. In this paper, we focus on the sampling and reconstruction situation when D is a sampling operator and H is identity.

To address this issue, there has been a growing amount of interest in recent years in applying results from the field of compressive sensing (CS) to imaging applications, an area known as compressive imaging. In 2006, the DSP group at Rice implemented the single-pixel camera which used the ideas of compressive sensing to efficiently capture images for the first time. The camera design reduces the required size, complexity, and cost of the photon detector array down to a single unit. The random CS measurements also enable a tradeoff between space and time during image acquisition[1]. Compressive sensing theory was initiated in 2006 by two ground breaking papers [2,3]. Since then, compressive sensing theory has been applied to problems in video processing [4,5], medical imaging [6], compressive imaging [7] and single image super-resolution[8].

2. COMPRESSIVE SENSING THEORY

In this section, we briefly review principles of compressive sensing. Readers seeking more detail are referred to the many papers on the subject found in the Rice University repository [9].

2.1 Principles of Compressive Sensing and Sparse Signals

Suppose we have an image represented by vector x with elements $x[n], n = 1, 2, ..., N$. The image can be represented in terms of the coefficients $\{s_i\}$ of an orthonormal basis expansion $x = \sum_{i=1}^{N} s_i \psi_i$ where $\{\psi_i\}_{i=1}^{N}$ are the $N \times 1$ basis vectors. We can concisely write the samples as $x = \Psi s$. The aim is to find a basis where the coefficient vector s is sparse or compressible (where only $K \ll N$ coefficients are nonzero). For example, natural images tend to be sparse in the discrete cosine transform (DCT) and wavelet bases.

The traditional image compression must compute and deal with a potentially large number of N transform coefficients $\{s_i\}$ even if only K coefficients are important. As an alternative, CS bypasses the sampling process and directly acquires a condensed representation using $M < N$ linear measurements between x and a collection of test functions $\{\phi_m\}_{m=1}^{M}$ as in $y[m] = \langle x, \phi_m \rangle$. We can write

$$y = \Phi x = \Phi \Psi s = \Theta s \tag{2}$$

where $\Theta = \Phi\Psi$ is a general $M \times N$ measurement matrix. The transformation from x to y is a dimensionality reduction and so loses information in general. In particular, since $M < N$, given y there are infinitely many x' such that $\Phi x' = y$. The magic of CS is that Φ can be designed such that sparse/compressible x can be recovered exactly/approximately from the measurements y .

To recover the image x from the random measurements y , the traditional techniques (e.g., inversion, least squares) do not work because of under-determined equation. However, compressive sensing have shown that if $M \geq 2K$ and A meets certain properties, then Eq.2 can be solved by using the l_1 optimization

$$\hat{s} = \arg\min \| s' \|_1 \text{ such that } \Phi\Psi s' = y \tag{3}$$

We can exactly reconstruct K -sparse vectors and closely approximate compressible vectors.

2.2 Restricted Isometry Condition

We cannot solve Eq.2 with any arbitrary $\Theta = \Phi\Psi$ until and unless Θ meets the Res. However, we can apply the compressive sensing framework if matrix Θ meets the Restricted Isometry Condition (RIC) condition [2,3]:

$$(1-\delta) \| z \|_2 \leq \| \Theta z \| \leq (1+\delta) \| z \|_2 \tag{5}$$

with parameters (z, δ) , where $\delta \in (0,1)$ for all z-sparse vectors z . In words, the RIC ensures that a measurement matrix will be valid if every possible set of z columns of Θ forms an approximate orthogonal set. In effect, we want the sensing matrix Φ to be as incoherent to the basis Ψ as possible. Only this condition is satisfied, we can use compressive sensing framework to Eq.2 and Eq.3.

2.3 Selection of Φ and Ψ

Examples of matrices for the sensing matrix Φ that may meet RIC include Gaussian random matrix (where the entries are independently sampled from a normal distribution), Bernoulli matrix (binary matrices drawn from a Bernoulli distribution), identity matrix, the noiselets, Walsh-Hadamard matrix, and the spike basis. The choice of sparse transform matrices for Ψ include DCT matrix, Fourier matrix, wavelet matrix and the overcomplete learned dictionary using the K-SVD, etc. The RIP criterion says that the sensing and transform matrices should have the least coherency or the greatest incoherency between them to yield better result.

ISBN: 978-989-8533-38-8 © 2015

2.4 Signal Recovery by Greedy Reconstruction Algorithms

The l_1-minimization method guarantees that sparse signals can be stably recovered. The OMP iteratively selects a set of vectors from the measurement matrix $\Theta = \Phi\Psi$ that contains most of the energy of the measurement vector y .The selection at the each iteration is made based on inner products between the columns of Θ and a residual, which reflects the component of y that is orthogonal to the previously selected columns.

A modification to OMP called Regularized Orthogonal Matching Pursuit (ROMP) was proposed. ROMP is similar to the OMP algorithm, in that it approximates the transform coefficients on every iteration and then sorts them in non-increasing order. However, the main difference is that instead of only selecting the largest coefficient, ROMP selects the continuous sub-group of coefficients with the largest energy, with the restriction that the largest coefficient in the group cannot be more than twice as big as the smallest member. These coefficients are then added to a list of non-zero coefficients and a least-squares problem is then solved to find the best approximation for these non-zero coefficients. The approximation error is then computed based on the measured results and the algorithm iterates again.

3. EXPERIMENTAL RESULTS

We have carried out a series of experiments with the standard gray images of "Lena" (256×256). We carry out the following experiments to study the performance of the image sampling and reconstruction based on compressive sensing.

3.1 Comparison between Reconstructed Images by Taking Different Sensing Matrices

(a) (b) (c)

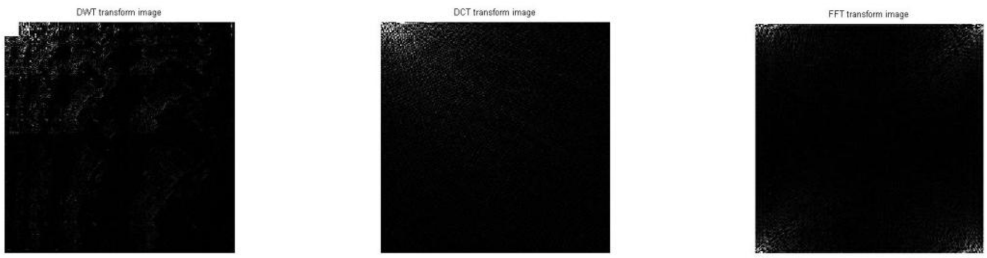
Figure 2. Subsampling rate $\alpha = 0.8$, taking wavelet transform matrix as Ψ : (a) Original image. (b) Taking Gaussian random matrix as Φ . (c) Taking Bernoulli matrix as Φ

We firstly performed a study on the influence of different sensing matrices on the quality of the output image. Fig.2 shows the reconstructed images obtained using the different sensing matrices. The PSNR is 31.5 and 31.3, respectively. The reconstructed images using the Gaussian random matrix and Bernoulli matrix yield the similar PSNR value.

3.2 Comparison between Reconstructed Images by Taking Different Sparse Transform Matrices

Figure 3. Subsampling rate $\alpha = 0.8$, taking Gaussian random matrix as Φ : (a) Taking Daubechies10 wavelet matrix as Ψ . (b) Taking DCT matrix as Ψ . (c) Taking Fourier matrix as Ψ

Next, we performed a study on the influence of different sparse transform matrices on the quality of the output image. Fig.3 shows the reconstructed images obtained using the different transform matrices. The PSNR is 31.5, 27.1and 26.6, respectively. Different transform matrix influence the reconstruction quality of the image. The reconstructed image using the Daubechies10 wavelet sparse transform matrix yields the highest PSNR value. We would like to use wavelets as Ψ because they are much better at sparsely representing images than bases such as DCT and Fourier (Fig.4).



(a) (b) (c)

Figure 4. Sparse transform images of "Lena": (a) DWT. (b) DCT. (c) FFT

3.3 Comparison between Reconstructed Images by Taking Different Recovery Algorithms

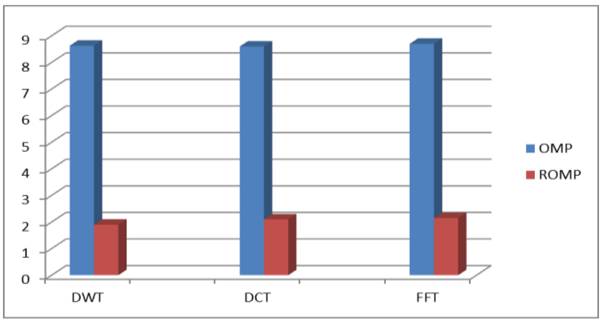


Figure 5. The time cost of reconstruction using OMP and ROMP recovery algorithms

For this purpose, we test time cost of OMP and ROMP with Gaussian random sensing matrix varying with different sparse transform matrices. Fig.5 clearly displays that the choice of sensing and transform matrices is independent with time cost and that ROMP runs faster than OMP algorithm to reconstruct images.

3.4 Comparison between Reconstructed Images by Taking Different Subsampling Rates

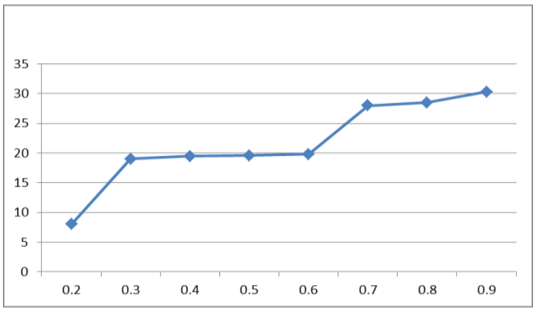


Figure 6. PSNR vary with subsampling rate (taking Gaussian random matrix as Φ and wavelet matrix as Ψ)

Fig.6 describes the curve of PSNR varying with subsampling rate. From visual comparisons, it has been observed that we reconstruct poor quality image with subsampling rate lower than 0.3.

4. CONCLUSION

In this paper, we have reviewed compressive sensing theory and carried out extensive simulations to study image sampling and reconstruction performance both visually and quantitatively. We have also studied the influence of various sensing and transform matrices, subsampling rate and reconstruction algorithms over the performance. It is concluded from the above studies that for better reconstruction, we should choose sensing and transform matrix to ensure the coherence between the two matrices should be minimum and increase the incoherence between them.

ACKNOWLEDGEMENT

This work was supported by The Natural Science Foundation of China (11271050, 11371183, 61403036), Science and Technology Development Foundation of CAEP (2012B0202024, 2013B0403068) and Beijing Center for Mathematics and Information Interdisciplinary Sciences (BCMIIS).

REFERENCES

[1] D. Takhar, et al., 2006, A New Compressive Imaging Camera Architecture Using Optical-domain Compression. *in Proc. of Computational Imaging IV. SPI*E. San Jose, CA.

[2] Candès, E., J. Romberg, and T. Tao. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489-509, 2006.

[3] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] R. Marcia and R. Willett. Compressive Coded Aperture Video Reconstruction. *in European Sig. Proc. Conf.*, Aug. 2008

[5] S.Y. Xiang, L. Cai, 2011. Scalable Video Coding with Compressive Sensing for Wireless Videocast. IEEE ICC proceeding, Kyoto, Japan.

[6] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[7] R. F. Marcia and R. M. Willett, 2008. Compressive Coded Aperture Superresolution Image Reconstruction. *IEEE ICASSP*, Las Vegas, America.

[8] B. Deka, M. K. R. Baruah. Single-Image Super-Resolution Using Compressive Sensing. *International Journal of Image Processing and Visual Communication*. Vol.1, Issue 4, pp.8-15, 2013

[9] "Rice University Compressive Sensing Resources website," 2009, http://www.dsp.ece.rice.edu/cs/.

[10] Richard G. Baraniuk. Compressive Sensing. *IEEE Signal Processing Magazine*, July 2007.

INTERACTIVE VISUALIZATION AND ANALYSIS OF EYE BIOMECHANICS

Iason Nikolas, Chris Papapavlou, Aris Lalos and Konstantinos Moustakas
University of Patras, Greece, Department of Electrical and Computer Engineering

ABSTRACT

Modeling realistic eye movements is an active area of research and can provide useful information related to brain and neurological diseases. To facilitate the early diagnosis and management of those diseases, both computational models of eye movements and visualization tools that will allow users to interact with those models are required. Accurate models are useful for simulating realistic movements, while visualization tools allow the extraction of meaningful information from the dynamic simulation. Several works in the literature focus on the eye movement simulation completely ignoring the information visualization part. In the context of this work a novel visualization tool is proposed, that illustrates the interaction between the biomedical eye movements and the extra-ocular muscles activation levels. This tool can be used as a test bed to understand and study how different neurological commands have a direct impact on the biomedical eye motion trajectory.

KEYWORDS

Information Visualization, Eye animation, Interaction, saccades, physics-based simulation.

1. INTRODUCTION

Recently, several experts in the fields of biomedical engineering and computer graphics, focus on modeling realistic eye movements. The movements of the human eye are divided in three categories: i) Saccades, ii) Vestibulo-ocular reflex and iii) Smooth pursuit. Saccades, are quick, simultaneous movements of the eye in the direction that serve to bring the visual target in the center of the retina. Vestibulo-ocular reflex, is the eye movement that stabilizes images on the eyes' retina during head movement, while smooth pursuit is the movement that follows smoothly a moving target. Saccades occur very frequently (up to 173,000 times per day) and they reflect the movements of visual attention (Remington 1980). More importantly, they provide information that can be useful for designing decision support systems that take clinical decisions related to brain and neurological disorders (Szewczyk-Krolikowski et al. 2014) (e.g., Parkinson, Alzheimer).

The need for novel visualization tools that will be able to effectively render the meaningful variables of the saccade movement are required, in order to help understanding these disorders. By visualizing the simulation results, analyzing and comparing them with measurements taken from healthy people or patients is essential for both clinical studies and modeling realistic eye movement.

1.1 Related Work

Most of the existing works of modeling eye movements, either focus on the smooth pursuit motion, completely ignoring the saccadic motions, generated by an oculomotor system, or they focus on the modeling of the saccadic movements, neglecting the information visualization part. The authors in (Lee et al. 2002) developed a simplified eye movement model, based on eye-tracking data and empirical methods that statistically evaluate how naturally generated saccades look like. Moreover, (Deng et al. 2005) in a more graphics like approach, propose the automated production of smooth eye movements using texture synthesis, while (Masuko & Hoshino 2007) in their work generate a composite head-eye movement for animation characters. In contrast with previously described works, the pioneering works of Robinson et al. (Robinson et al. 1969; Robinson 1964) provided a computational biomedical model of the human eye and the associated muscles, allowing the realistic simulation of saccadic movements. However, these works were limited in one dimensional movements. The authors of (Papapavlou & Moustakas 2014), provide a realistic biomechanical human eye model, incorporating anatomical details of the oculomotor system using the platform described in detail by (Delp et al. 2007) and (Sherman et al. 2011). Although, this model allows the dynamic simulation of realistic eye saccadic motions, based on the corresponding muscle activation levels, it completely ignores the visualization part, which as stated before is essential i) for giving the feedback that is needed to fine tune the models' parameters and ii) to study those movements when presented with neurological and muscle disorders in the system.

1.2 Contribution

To the best of our knowledge, this is the first framework that can be used for visualization of realistic saccadic eye movements. It is expected to be very helpful in the analysis of the simulation results and in the formulation of novel biomarkers that can be estimated unobtrusively from the behavioral profile of the saccades. More precisely the contribution of our work is the following: i) animation of the eyeball and its muscles based on the realistic simulation, ii) interaction with the eye model and the ability to start a simulation based on a point in 3D space and specific parameters, iii) information visualization in 2D plot diagrams of the geometrical characteristics and activation levels of the muscles during a simulation. Comparison between two or more simulations can be performed by drawing the results in the same plot diagram. iv) 3D representation of trajectories during the simulation and v) color mapping on muscles based on their activations, vi) an easy to use interface to modify parameters of the system.

2. EYE BIOMECHANICS

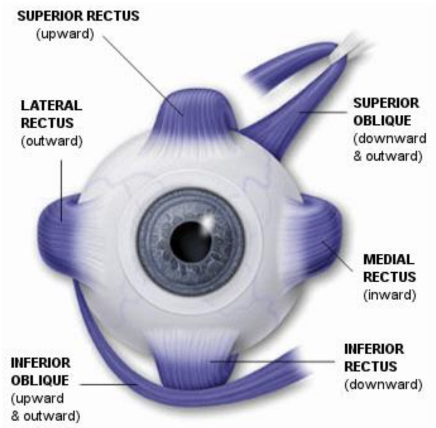


Figure 1. Front view of the right eyeball. The extra-ocular muscles are also presented

A right eyeball front view with extra-ocular muscles is presented in Figure 1. The human eye is represented as a perfect symmetrical sphere which is able to rotate around its center. The mean value of the eyes' diameter for adults is 24mm and the mass is approximately 7.5grams, with small deviations between individuals. The eye movement is generated by six extra-ocular muscles, known as: i) superior rectus, ii) inferior rectus, iii) lateral rectus, iv) medial rectus, v) superior oblique and vi) inferior oblique. Apart from inferior oblique, all the others have a common origin, which is a ring of fibrous tissue behind the eyeball and is called Annulus of Zinn. The annulus of Zinn is not exactly behind the eyeball and this asymmetry is also taken into account in the considered model. The four recti muscles (i)-(iv) have straight forward functionality in comparison with the other two oblique muscles (v)-(vi). All six muscles can be grouped as three agonist-antagonist muscle pairs. Superior and Inferior rectus are responsible for upward/downward movement. Lateral and Medial rectus are responsible for right/left movement. Finally the two oblique muscles cannot be assigned to a specific rotation axis since their insertion points enable them to contribute virtually in all the movements. At every fixation point, the muscles are equilibrating the elastic restoring forces of the passive tissues in order to keep the eye steady. Before and after the saccade occurs, the initial and final orientations are determined through the steady-state activations, which correspond to the state where the eyeball is stiff at its position. The activations during the transition from the initial to the final state, are called transient muscle activations. Those activations are different from a discrete step response. According to (Robinson et al. 1969) there is no active control during the movement. As a result the eyeball system is considered as an open-loop system and muscle activations are already determined as functions of time, by the brain, before movement begins. The response times regarding saccades differentiates between humans. This behavior is affected by many factors such as mood, fatigue, etc. With appropriate configuration of the signal parameters, the user can modify the produced saccades in order to simulate the variations in saccade response.

3. INFORMATION VISUALIZATION AND INTERACTION

Table 1. Editable Parameters

Physiological / Behavioral Parameters	
Mass	eye mass
Radius	eye radius
Range	field of view angle
Stiffness	passive stiffness
Viscosity	viscosity
Max force	the maximum force produced by a muscle
Fiber Length	the optimal fiber length
Tendon Slack Length	the tendon rest length at rest state
Activation	constant activation level of each muscle

Based on the previously described model, we developed a novel visualization tool that illustrates the interaction between the eye movements and the extra ocular muscles activation levels. The implemented tool allows the presentation of the information that is generated by the saccadic eye simulation, in a way that is easily perceived by the clinicians and researchers. Auxiliary data are also provided to help the user at the first glance get the overview of the simulation results. Then the user is capable of visualizing the results with the level of detail that is required in order to come to a conclusion. To that end a graphical user interface (GUI) was implemented, allowing the user: i) to modify, apply and save all the eye and muscle parameters (e. g., stiffness, tendon slack length) and ii) to set the activation levels of the involved muscles in order to perform a forward simulation.

In practice, muscles are wrapped around each eye. To illustrate this feature we use two attachment points and two to four intermediate. These points are connected by muscles, which are represented as cylinders. Then the eyes, represented as spheres, are attached at those points. Several other intermediate points can be generated using splines, in order to construct a smoother muscle envelop. For the illustration of the saccade movement simulation results, both 2D plots and an animation 3D Viewport are implemented. An illustrative example is presented in Figure 2. Rotation results can be visualized through 2D plots, trajectories and animation. Rotations and activations plots have the same time axis and are colored differently for each muscle. Trajectories of the eyes are projected either on a plane in front of the eyes, or in the pupil of each eye. The animation of the current simulation provides information for the rotation of each eye, the muscle activation and the force applied at each time instance. To be more specific, the activation levels of each muscle are mapped to a color gradient that ranges from blue to red, while the stress of each muscle is visualized based on the force that the muscle applies at each time instant, by changing its radius from a minimum to a maximum value.

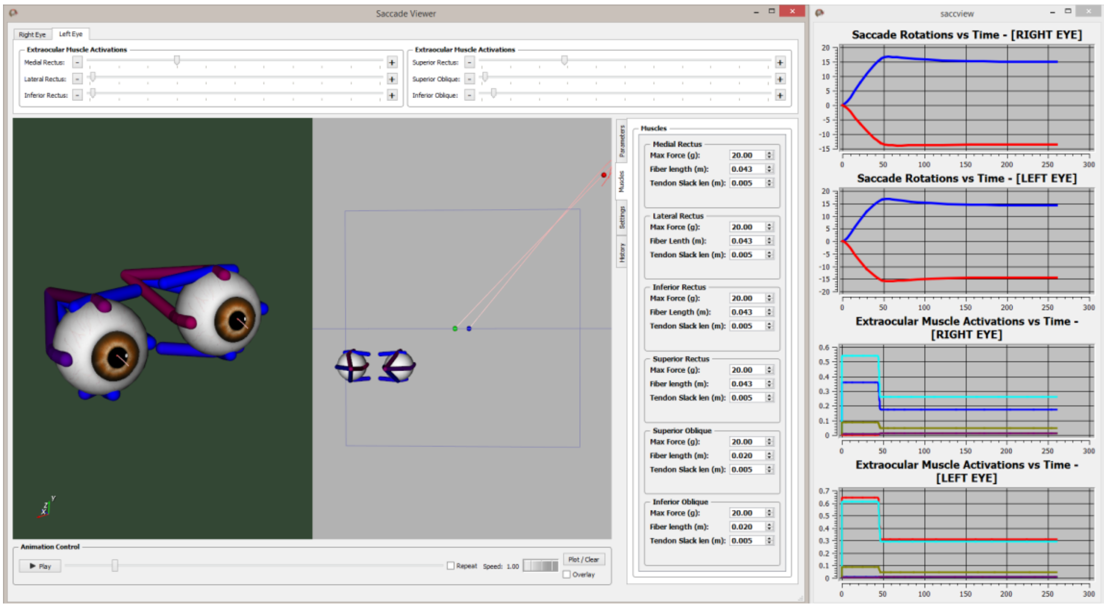


Figure 2. Frameworks' GUI has two tabs at the top of the main window that enables the user to set the activation level of each eye muscle and produce the forward simulation. The tabs on the right makes all setting editable. On the left-center there are two interactive 3D Viewports that gives the user the ability to carefully preview the results of the simulation. To be more specific, user can see the front and the back of the eyes, while at the same time trajectories are displayed at the pupil and the plane in front of the eyes. At the right of this figure we can see a window that contains the 2D plots of the muscle activation levels and the eye rotations that produced during the simulation. Finally, at the bottom of the main window the user can manipulate the animation time and speed.

Considerable attention was given to the interaction with the user. First of all, it should be noted that the GUI is fully functional and enables the user to fully customize both physiological and behavioral parameters. Physiological parameters are fixed and depend to the eye anatomy of a specific patient. While the extraction of behavioral parameters during different saccades, could allow a clinician to extract useful biomarkers that can be used for diagnosis of neurological diseases. A synopsis of the framework capabilities and parameters, is shown in Table 1.

The main interaction with the user takes place in the 3D Viewport which is separated into two sub-windows as shown in the left side of the Figure 2. On the left window, the cameras' focal point is located between the eyes, while in the right window the cameras' focal point is located in the center of the plane in front of the eyes. By clicking and moving the mouse we can rotate each camera on a sphere around each focal point. In addition, the user is also capable of i) panning, rolling and resetting the camera to each initial state ii) zooming in and out of the focal point iii) setting multiple saccade eye movements with the same initial orientation. The last feature facilitates the comparison between different saccades. Simulation results Parameters that can be visualized include muscle forces, torques, activation signals, muscle state (length, velocity), trajectories and geometric deformation. Researchers are usually interested in geometrical parameters and the activation levels of the muscles. In the 3D Viewport the animation is playing while at the same time corresponding sliders are moving and the color of the muscles is changing according to each activation level. An overview of our visualization tools' GUI is shown in Figure 2. In the main 3D Viewport the animation and the trajectories are drawn to give to the users a quick overview of the simulation results. The activations that were produced and the rotations of the eye are plotted as shown in the right side of Figure 2 and Figure 3. It is really important for the users to be able to compare two or more simulation results. An example of simulation comparison can be seen in Figure 3. The same plot displays rotations of two different simulations that have the same initial and target points but different stiffness and viscosity parameters. Through this diagram, the user can perceive the effect of stiffness and viscosity of the saccadic eye motion. In the same way comparison can be made between activations of different muscles to see how they responded to each situation. To conclude, by using this visualization tool, the researchers can easily make assumptions about how each parameter of the model affects the saccadic eye movement.

4. CONCLUSIONS

In this paper a novel visualization tool that allows the monitoring of simulation results has been proposed that visualizes eye movements and the respective muscle activation levels. It is worth mentioning, that the proposed framework can play a major role in the treatment of diseases related to neurological disorders, by visually assisting researchers. Future research will focus on simulating and visualizing neural signals generated from a brain simulator, described in (Blenkinsop et al. 2014). The neural signals that are related to a specific saccadic eye movement will be then passed to the presented simulation engine (forward simulation) in order to evaluate both neurological and muscle characteristics.

This integration is of significant clinical importance for Parkinson's disease (PD). By simulating the coupled brain-eye biomechanics system we aspire to model and predict specific features of high clinical significance for PD like saccadic latency and motion trajectory overshoot. The latter are already analyzed in clinical practice using specialized devices called saccadometers. Finally, the clinicians will be able, using the proposed visualization framework, to dig into the detail of both the simulations and the experimental saccadometry data.

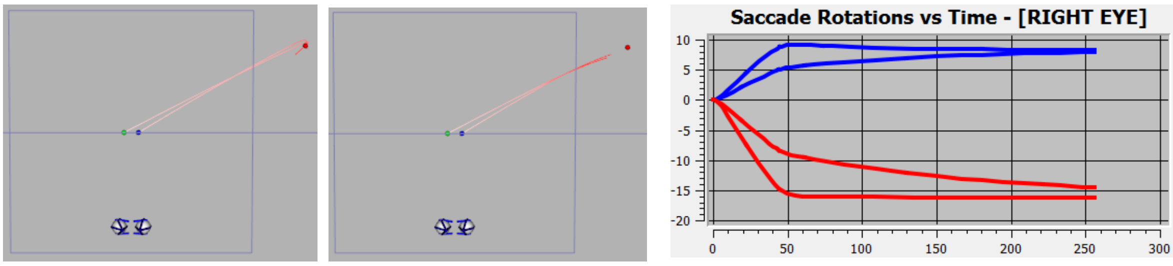


Figure 3. Comparison of two simulations with the same initial and target point, but with different stiffness and viscosity parameters. Each one of the respective trajectories (left) can be drawn by selecting each simulation from history tab, while both rotations are plotted in the same diagram (right).

ACKNOWLEDGEMENTS

This work is partially funded by the EC FP7 project NoTremor: Virtual, Physiological and Computational Neuromuscular Models for the Predictive Treatment of Parkinson's disease, Grant Agreement No. 610391.

REFERENCES

Blenkinsop, A. et al., 2014. A computational framework for describing the saccadic eye movement system of the Parkinsonian digital patient. *Virtual Physiological Human Conference 2014*, (8), pp.9–11.

Delp, S.L. et al., 2007. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11), pp.1940–1950.

Deng, Z., Lewis, J.P. & Neumann, U., 2005. Automated eye motion using texture synthesis. *Computer Graphics and Applications, IEEE*, 25(April), pp.24–30.

Masuko, S. & Hoshino, J., 2007. Head-eye animation corresponding to a conversation for CG characters. *Computer Graphics Forum*, 26(3), pp.303–312.

Papapavlou, C. & Moustakas, K., 2014. Physics-based modelling and animation of saccadic eye movement. *22nd International Conference on Computer Graphics, Visualization and Computer Vision, WSCG 2014, Plzen, June 2014*.

Remington, R.W., 1980. Attention and saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 6(4), pp.726–744.

Robinson, D. et al., 1969. Mechanical components of human eye movements. *Journal of Applied Physiology*, 26(5), p.548. Available at: http://jap.physiology.org/content/26/5/548.full.pdf.

Sherman, M. a., Seth, A. & Delp, S.L., 2011. Simbody: Multibody dynamics for biomedical research. *Procedia IUTAM*, 2, pp.241–261. Available at: http://dx.doi.org/10.1016/j.piutam.2011.04.023.

Szewczyk-Krolikowski, K. et al., 2014. Functional connectivity in the basal ganglia network differentiates PD patients from controls. *Neurology*, 83, pp.208–214.

ISBN: 978-989-8533-38-8 © 2015

ROCK FALL SIMULATION USING 3D MODELS RECONSTRUCTED FROM MULTIPLE-VIEW IMAGES

Shogo Atsumi, Itaru Kitahara and Yuichi Ohta
Department of Intelligent Interaction Technologies, University of Tsukuba
Tennoudai 1-1-1, Tsukuba Science City, Ibaraki, Japan

ABSTRACT

This paper proposes a method of rock fall simulation using physics engine and 3D models reconstructed by Image-Based Modeling technique. The shape of rocks is important to estimate the accurate way of leap and route of falling motion in 3D simulation. However, in most current researches, although the 3D model of slope is well accurately scanned using a laser scanner, it is not sufficient to estimate the complex behavior of falling rocks due to modeling the shape of rocks is too simple such as sphere or cube. Our proposed method generates 3D models in more accurate shape using Image-Based Modeling technique. The estimated scales are precisely reconstructed as compared with the real measured ones. We evaluate our method in some different places and conduct on rock fall simulation. As the result, by applying physics engine to simulation of falling motion, we confirm that the shape of falling rocks improves the estimation of falling route.

KEYWORDS

Rock fall simulation, Multi-view images, 3D reconstruction, Structure from motion, Physics engine.

1. INTRODUCTION

It is necessary to estimate the route, height of leap and velocity of the rock for prevent rock fall. However, rock fall is a complex phenomenon affected by position, size, or shape of rocks and angle of slope. Researches about estimation of falling motion of rocks are actively conducted (rock fall simulation) (Nguyen H.T, 2011). Since the shape of a slope is treated as the 2D cross-section in these most simulations, it is difficult to estimate the 3D route, height of leap and velocity of the rock accurately.

Some researches attempts to expand the dimension from 2D to 3D. Although such simulation can acquire shape of slopes using GIS (Charalambous, 2008), it is still difficult to measure the shape enough for precise rock fall simulation. Moreover, in order to accurately estimate the collision with rocks and slope, the shape of rocks should be represented accurately as possible. By using a laser scanner, shape of slopes and rocks can be accurately measured. However, it is not feasible to bring and install such heavy and delicate equipment on a steep slope where fall rocks exist.

As Figure 1 shows, we propose a method of rock fall simulation using the 3D model of rock and slope reconstructed by merging multiple-view images (Image-Based Modeling). Since our proposed method needs only mobile camera to generate 3D model, it is possible to realize more practical rock fall simulation. In addition, using our system, inspectors can virtually confirm how the rock falls on the slope, since textured 3D models are reconstructed.

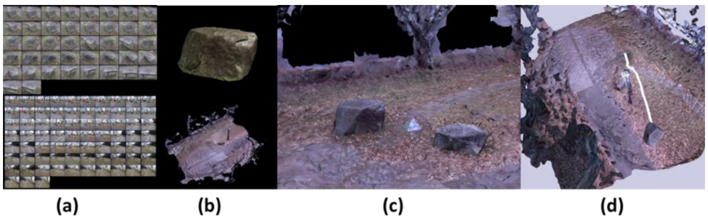


Figure 1. Overview of our proposed method ((a): Multiple view images of rock and slope, (b): Reconstructed 3D models of rock and slope, (c): Overlaid rocks on a slope, (d): Rock fall simulation using 3D models).

2. RELATED WORKS

Christen M. et al (Christen M. et al, 2012) conduct on 3D rock fall simulation using slope shape scanned by a laser scanner. However, laser scanner is too heavy to be installed on a slope. Some previous methods approximate the shape using simple primitive such as sphere or cube. Unfortunately, it is difficult to estimate the complex behavior of falling rocks.

Structure from Motion (SfM) (Tomasi T., 1992) is one of the most popular 3D reconstruction methods using multiple-view images To apply Patch-based Multi-View Stereo (PMVS) (Furukawa Y., 2009) improving the density of reconstructed 3D data and Poisson Surface Reconstruction (PSR) (Kazhdan M. et al, 2006) generates 3D models with surface information.

In order to realize fall rock simulation, it is necessary to merge multiple 3D models of rock and slope, which are individually reconstructed. Iterative Closest Point (ICP) (Besl P.J., 1992) is well known as the alignment method of 3D models. Although, it is intended for 3D models that have comparable spatial resolution, the definitely of our target 3D models is completely different. (i.e., the 3D rock model has high spatial resolution. the 3D slope model has low resolution.) On the other hand, since our method generates textured 3D model, it is possible to merge them using the correspondence. Scale-Invariant Feature Descriptor (SIFT) (Lowe D.G., 1999) is well known method to detect correspondence, which is robust to the appearance variation. However, when there is perspective deformation in the images, matching accuracy of SIFT is reduced. On the other hand, in our method, by rendering an image of rock that looks similar in a slope image, it is possible to estimate the corresponding points and estimate the 3D rigid transformation.

3. RECONSTRUCTING 3D MODELS OF ROCK AND SLOPE

3.1 Reconstructing 3D Models of Rock and Slop

Figure 2 shows processes to reconstruct 3D model of a rock. We capture the target rock with getting closer and going around (a). We also record the orientation of capturing camera. The adequate number of images to reconstruct fine shape 3D model is thirty at least, it depends on size and complexity of the rock, though. A slope around the rock is also captured using multiple cameras. Although SfM can generate a sparse point cloud data (b), it is possible to increase the density of the point cloud by applying PMVS (c). We reconstruct surface information to execute the physical simulation by applying PSR (d). Laplacian smoothing (Nealen A. et al, 2006) is applied to the reconstructed 3D slope to reduce rough and uneven shape.

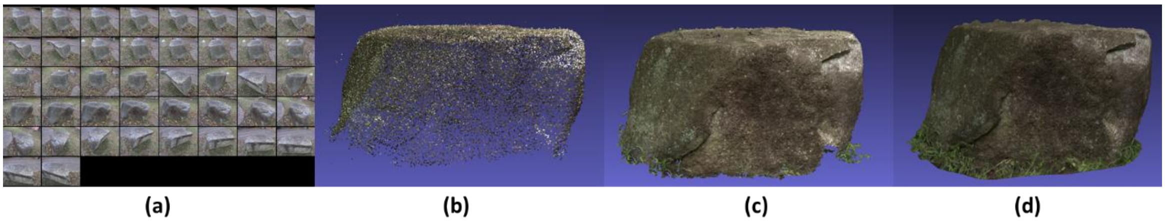


Figure 2. Processing to reconstruct a 3D model from multiple-view images.

3.2 Correspondence between Rock and Slope 3D Models

In our method, we have to solve two problems to estimate 3D rigid transformation matrix. One is the significant difference in their spatial resolution; the other is the difference in the appearance caused by perspective projection. We solve the first problem to find out corresponding points through 2D image processing instead of directly finding them on the 3D models. The second problem can be solved by generating an appropriate appearance using the 3D rock model that has fine texture and shape information.

As shown in Figure 3, if we set a virtual camera at the same place of a camera capturing the target slope and render the 3D rock model, it is possible to make virtual appearance of the rock similar as the real appearance in the slope image. As the result, it is possible to easily detect corresponding points using image feature such as SIFT. As the initial setting of a virtual camera pose, we use positional information given by GPS (Global Positioning System). The orientation information given by an electronic compass is also used as the initial orientation. Measurement error of GPS is about 10 m on average and the orientation information obtained by an electronic compass has a few degrees of error (Yun X. et al, 1999). Thus, we shit and swing the virtual camera within the error distribution. We estimate the camera pose assuming that a virtual camera at an appropriate pose gives the most number of the matching points.

Figure 3. An image of slope (left), A rendered image of 3D rock model (right)

3.3 Merging Reconstructed 3D Models

If we have four corresponding points between the 3D rock model and the 3D slope model, it is possible to calculate the 3D rigid transformation matrix. 3D coordinate of the corresponding points can be estimated as an intersection of a ray of a target point and 3D slope model. When corresponding points of 3D rock model are $\mathrm{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N\}$ and the corresponding points of 3D slope model are $\mathrm{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N\}$, X and Y are expressed as follow by using a transformation matrix $\boldsymbol{H}$.

$$\boldsymbol{y}_i = \boldsymbol{H}\boldsymbol{x}_i \ (\mathrm{i} = 1, \cdots \mathrm{N}), \boldsymbol{H} = \begin{bmatrix} \boldsymbol{SR} & \boldsymbol{t} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \quad (1)$$

The Matrix $\boldsymbol{S}$ is expressed as $\boldsymbol{S} = s\boldsymbol{I}$ by using a unit matrix $\boldsymbol{I}$ and the scaling factor s. It is possible to obtain the factor s by comparing a distance between two points in the corresponding points. After obtaining a center of gravity of the point X and points Y, a translation vector $\boldsymbol{t}$ is calculating by taking the difference between the Euclidean distance of the two points. A rotation matrix $\boldsymbol{R}$ is obtained by singular value decomposition (Kanatani K., 1994).

4. EVALUATION OF ACCURACY OF RECONSTRUCTED 3D MODELS

4.1 Accuracy of Reconstructed 3D Models

In order to estimate a scale parameter of the 3D reconstruction, a square pyramid, which the size is already known, is set on the environment and captured with a target rock. This square pyramid is composed of four equilateral triangles (one side is 0.5m). We attach textures to get many image features on each surface to improve the accuracy of the 3D reconstruction.

We conducted on experiments to confirm the accuracy of generated 3D model. We capture images using RICOH CX2 with 3456 pixels ×2592 pixels resolution as shown in Figure 5. As Figure 4 (left) shows, the square pyramid is located near Rock A and Rock B. The number of images capturing Rock-A is 52 and Rock-B is 32. Figure 4 (right) shows the position where we measure the actual size. As the result, the average error is 8.29×10^{-3}m and the max error is 1.50×10^{-2}m. We confirm that the 3D model can be reconstructed accurately.

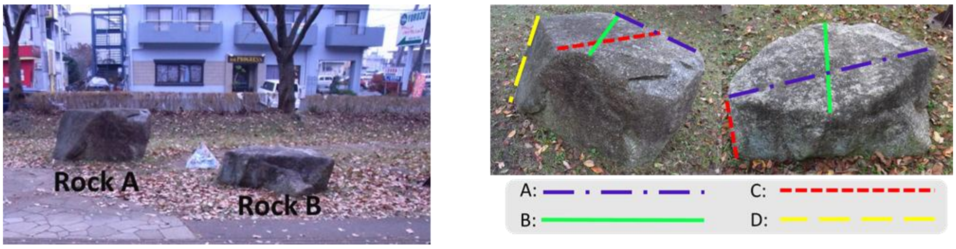
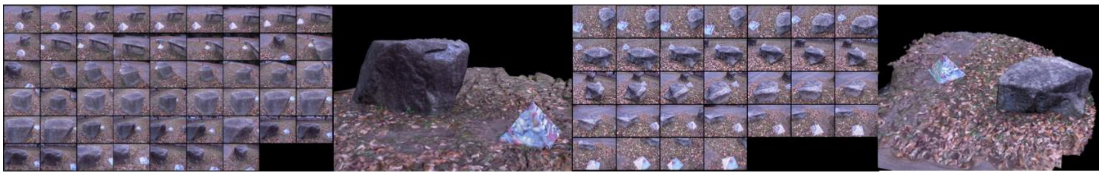


Figure 4. Spatial relationship among rocks and a quadrangular pyramid (left), Positions we measure the actual size (right).



(a) (b)

Figure 5. Captured multiple-view images (left), Reconstructed 3D model (right) ((a): Rock-A, (b): Rock-B).

4.2 Accuracy of 3D Models Alignment

We also capture images of the environment (162 images with 3456 pixels × 2592 pixels resolution using RICOH CX2), reconstruct the 3D model and merge the 3D models with the 3D models of Rock-A and Rock-B. The rigid transformation matrix is calculated using the obtained corresponding points. The number of the corresponding points of Rock-A and Rock-B is 52 and 60, respectively. The result, we confirm 3D models of Rock-A and Rock-B is precisely merged with a small error (7.35cm and 10.5cm on average, respectively).

5. EXPERIEMENT AT LANDSLIDE SITES, ROCK FALL SIMULATION

We capture multiple images of rocks and slopes at actual landslide site (Syosenkyo, Kofu, Japan) to confirm the practicality of our proposed method. A rock were captured at 29 viewpoints using CANON EOS 5D MarkII with 5616 pixels × 3744 pixels resolution. A slope was capture using SONY FDR-AX1 with 3840 pixels × 2160 pixels resolution and 169 frames were used to reconstruct 3D slope model. Figure 6 shows the images captured at the site and the reconstructed 3D models.

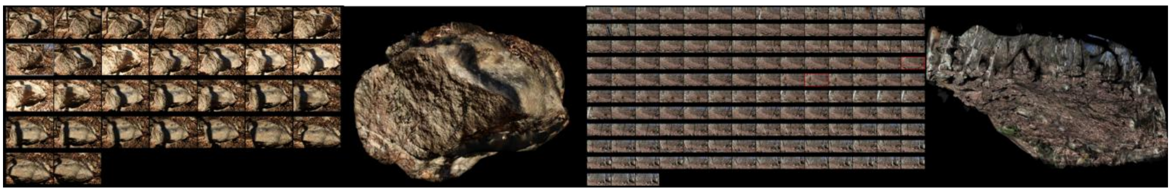


Figure 6. Multiple-view Images (left), Reconstructed 3D model (right) at Syosenkyo (left: rock, right: slope)

Bullet physics (Bullet Physics Library, 2014) is used for rock fall simulation. The result of the simulation is drawn using OpenGL. We set the density of 3D space as 2500km/m^3, coefficient of dynamic friction to 0.35 and coefficient of restitution to 0.4 (Japan Road Association, 2000). In the simulation, the mass of 3D slope model, which is static rigid body, is zero. 3D rock model is dynamic rigid body. We use a PC (DELL Studio XPS, CPU:3.07[GHz](Intel Core i7), GPU:NVIDIA GeForce GTS 240) to reconstruct 3D model and simulate rock fall.

Figure 7 shows a result of rock fall simulation by using 3D models of Syosenkyo. In order to confirm the effectives to reconstruct accurate 3D shape, we also simulated rock fall by approximating the shape as sphere and cube. The each trajectory of the center of gravity is drawn with white, red and green line. As the result, we confirm our 3D rock model's motion (white line) is completely different from the other shape, sphere (red line) and cube (blue line). The shape of rock is important parameter to estimate the fall path.

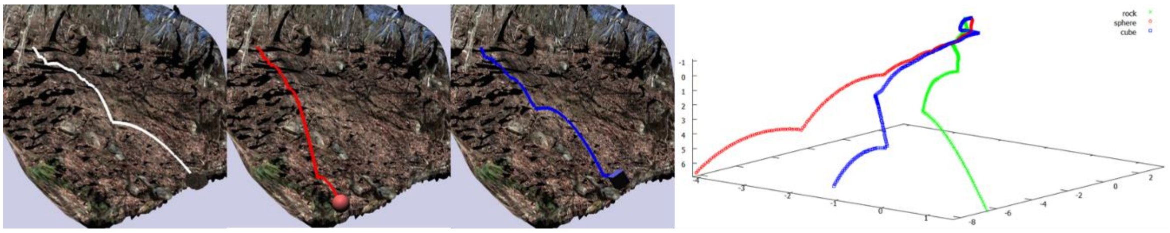


Figure 7. A result of rock fall simulation by using CG objects (From left to right: 3D rock model, Sphere, Cube, A plot of 3D trajectories (green: 3D rock model, red: sphere, blue: cube))

6. CONCLUSION

This paper proposed the method to simulate rock fall by using 3D model reconstructed from multiple-view images. By detecting the correspondence between the models of rocks and a slope utilizing free-viewpoint image technique, we can merge two 3D models. In addition, we realized a rock fall simulation without modeling a motion of flying-jumping, sliding and tumbling. In the experiment, we confirmed that the shape of generated 3D models and the alignment of 3D rock and slope models are enough accurate for rock fall simulation.

REFERENCES

Nguyen, H.T. et al, 2011. Use of terrestrial laser scanning for engineering geological applications on Volcanic rock slopes an example from Madeira Island (Portugal). *Natural Hazards and Earth System Sciences*, Vol.11, pp.807-817.

Charalambous, S. and Sakellariou, M., 2008. GIS-Based Rockfall Hazard Assessment in Support of Decision Making. Laboratory of Structural Mechanics School of Rural and Surveying Engineering, *National Technical University of Athens.* Athens, Greece.

Christen,M. et al, 2012. Integral hazard management using a unified software environment: numerical simulation tool" RAMMS" for gravitational natural hazards. *12th Congress INTERPRAEVENT*, Grenoble, France, pp. 23-26.

Chen, Y., Medioni, G., 1992. Object modeling by registration of multiple range images. *Image and Vision Computing*, Vol.10, No.3, pp.145-155.

Tomasi, C., Kanade, T., 1992. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, Vol.9, No. 2, pp. 137-154.

Furukawa, Y., Ponce, J., 2009. Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362-1376.

Kazhdan, M., et al, 2006. Poisson surface reconstruction. *Symposium on Geometry processing*, pp.61-70.

Besl, P.J., Mckay, N.D., 1992. A Method for Registration of 3D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.2, pp.239-256.

Lowe D.G., 1999. Object recognition from local scale invariant features. *The proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, pp.1150-1157.

Nealen, A., 2006. Laplacian mesh optimization. *ACM SIGGRAPH*, pp. 381–389.

Yun, X., et al, 1999. Testing and evaluation of an integrated GPS/INS system for small AUV navigation. *Oceanic Engineering, IEEE Journa*l, Vol.24, No.3, pp.396-404.

Kanatani, K., 1994. Analysis of 3-D rotation fitting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 5, pp. 543–549.

Bullet Physics Library, 2014, http://bulletphysics.org/wordpress/

Japan Road Association, 2000. *Handbook of preventatives against rock falls*. Maruzen, Japan. (in Japanese)

3D FOOT SCANNING IN FLEXIBLE MANUFACTURING: COMPARISATION AND DEFORMATION SOLUTION

Bita Ture Savadkoohi[1] and Raffaele De Amicis[2]
[1]*Seraj Higher Education Institute , Department of Computer and Electrical Engineering*
No.283, Bahar Cross Road, Monajem Street, Tabriz, Iran
[2]*Fondazione GraphiTech, Via Alla Cascata, 56/c 38123 Povo, Trento, Italy*

ABSTRACT

One of the new emerging areas of ubiquitous manufacturing is “Mass Customization”. A new production technology (also called 3D non-contact optical scanning) has been developed that can revolutionize the way how mass customization products are delivered by fast, flexible, and cost-effective production directly from 3D foot data to satisfy the fit and comfort requirement of costumers. In this paper in order to produce the right fit and comfort footwear, an algorithm for cutting 3D triangle mesh to several sections from heel toward toe is proposed. Then the area of each contour is calculated and searched in shoe last data base for finding the similar shoe last’s contour. If the area and shape of provided foot’s contour and shoe last’s contour are the same, that one is chosen as the best fit, otherwise the shoe last’s contour which has the same or near area by different shape is deformed based on the foot’s contour.

KEYWORDS

Mass customization, footwear, shoe last.

1. INTRODUCTION

The design of new shoes starts with the design of the new shoe last. In such industry, the shoe last which is a mechanical form should be designed rapidly from the individual foot model. A meaningful way to evaluate footwear compatibility would be to determine the dimensional difference between foot and shoe last. An approach to computerize footwear fit is proposed in this paper. In such system shoe last can be scanned in the same way as feet and stored in the shoe last data base. All 3D models in this paper are reconstructed with triangle mesh and all mesh models are assumed to be oriented, manifold, connected and aligned from heel to toe [1] [2]. To full fill the task an algorithm for cutting 3D models to several sections from heel to toe is proposed. Then contours which are one of the critical design features to produce the right fit and comfort should be extracted. In such system shoe lasts are cut to several sections and contours are extracted and stored in the shoe last data bases. Then 3D customer’s foot is cut in the same way and searched in shoe last data base for finding similar contour. If the contour with same area and same shape is found, that one is chosen as best fit. Otherwise the shoe last’s contour with same or near area is deformed base on the customers foot’s contour. So that best fit can be obtained relatively automatically and quickly. This paper is structured as it flows: In Section 2 an algorithm for cutting the 3D model to several sections from heel to toe is introduced, while in Section 3 deformation method is described. Finally, the conclusion remarks are gathered in Section 4.

2. CONTOUR EXTRACTION FROM 3D MODEL

The similarity search algorithm is based on the cutting 3D models into several sections from heel toward toe. Then the area of each section (available contour) is calculated and compared with the area of equal sections in shoe last data base. Let T={T_1 ,….,T_n }$(T_i \subset R^3)$ be a set of "triangles of mesh", V={V_1,…,V_n}(V_i=(x_i, y_i, z_i)) and E ={E_1, E_2,…., E_n} be set of “**V**ertices” and **E**dges associated to triangle mesh, N_v , N_e , N_t be **N**ew

ISBN: 978-989-8533-38-8 © 2015

Vertices, **N**ew **E**dges and **N**ew **T**riangle, respectively established by cutting plane, $V_c=\{V_1,\ldots,V_{nvc}\}(V_i=(x_i, y_i, z_i))$ and $E_c=\{E_1, E_2,\ldots, E_{nec}\}$ be set of "**V**ertices" and **E**dges associated to **C**ontour after cutting mesh with intersection plane, C_{gc} be **C**enter of **G**ravity of **C**ontour, N_{tc} be the **N**umber of **T**riangles of **C**ontour that associated with edges of contour (E_c) and center of gravity of contour (C_{gc}) and A_{tci} be the **A**rea of each associated **T**riangle in **C**ontour. As shown in Figure 1.a, E_1 is a common edge between two triangles T_1 and T_2, V_2 is the opposite vertex of E_1 in T_1, V_4 the opposite vertex of E_1 in T_2, E_3 is previous edge of E_1 in T_1 and E_4 is next edge of E_1 in T_2. The main steps of the proposed cutting algorithm are:

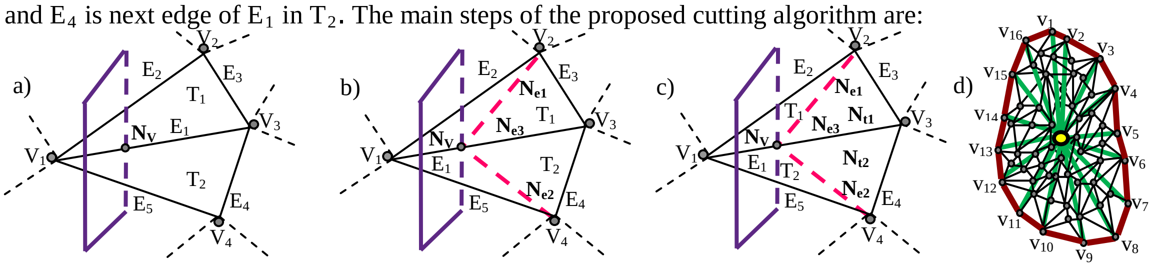


Figure 1. a, b and c are steps for cutting the shape to several sections. d) The illustrated contour after cutting mesh

Step 1. Find the intersection of the cutting plane with the edge (E_1) of a triangle and create new vertex (N_v), and find the next edge (E_4) and the previous edge (E_3) of the current edge (E_1), see Figure 1.a.
Step 2. Build a new edges (N_{e1} and N_{e2}) between the intersection point (N_v) and the opposite vertices (V_2 and V_4) of the current edges, see Figure 1.b.
Step 3. Build a new triangle between the new edges, see Figure 1.c, (e.g, $N_{t1}=(N_{e1}, N_{e3}, E_3)$ and $N_{t2}=(N_{e2}, N_{e3}, E_4)$
Step 4. Update the triangles and vertices of the current edge, see Figure 1.c, (e.g, $T_1=(E_1,N_{e1},E_2)$, $T_2=(E_1,N_{e2},E_5)$ and $E_1=(V_1,N_V)$) and add the new vertex, edge and triangle to the list (e.g, add N_v to vertex list, add N_{e1},N_{e2} and N_{e3} to edge list and add N_{t1} and N_{t2} to triangle list).
The output of our algorithm is a set of vertices and edges related to each contour as is illustrated in Figure1.d.
The main steps for calculation area of each contour are described in following steps:
Step 1. Calculate center of mass of the contour as follows:

$$x_c=\frac{1}{N_{tc}}\sum_{i=0}^{N_{tc}}x_i\ ,\quad y_c=\frac{1}{N_{tc}}\sum_{i=0}^{N_{tc}}y_i\quad \text{and}\quad z_c=\frac{1}{N_{tc}}\sum_{i=0}^{N_{tc}}z_i \tag{1}$$

So the point $C_{gc}(x_c, y_c, z_c)$ is the **C**enter of **G**ravity of **C**ontour.
Step 2. For calculating the area of each contour, we divide the vertices of contours edge and center of gravity of contour in triangles, see Figure 1.d. The area of each contour can be evaluated as follows:

$$A=\sum_{i=0}^{N_{tc}}A_{tci} \tag{2}$$

Step 3. Search in shoe last data base for finding shoe last's contour with same shape and area. Otherwise deform the shape of shoe last's contour based on foot's contour to evaluate footwear fit.

3. DEFORMATION OF CONTOUR

The research field of geometry processing studies is the entire life of a shape. A shape is born by geometry acquisition (e.g. scanning a physical object) and the life of a shape ends with consumption. Throughout the other stages of this geometry processing, a shape is analyzed and manipulated. One of those manipulations stages is shape deformation, where a shape's geometry is changed to achieve a specific goal. Surveys on space deformation technique have been presented in [3] and a general introduction to shape deformation technique is provided in [4]. For deformation shoe last's counter with respect to consumer foot's contour, we use deformation method proposed by He et al.[5] which is based on Feature Sensitive (FS) metric. FS deformation method that is an extended Laplacian for deformation, better preserves mesh details (especially feature) than Laplacian methods, but the time complexity of this method is similar to existing linear Laplacian methods.

3.1 FS Laplacian Operator

Let M= (V, E, F) be a given triangle mesh, where V be set of vertices, E be set of edges and F be set of faces, $\hat{M} = (\hat{V}, \hat{E}, \hat{F})$ be coordinates relates to dual mesh, w_{ij} be the edge weight in edge(i,j), $N(i) = \{j | (i, j) \in E\}$ be neighbors of i, α_{ij} and β_{ij} be two angle opposite to edge(i,j), see Figure 2.a.

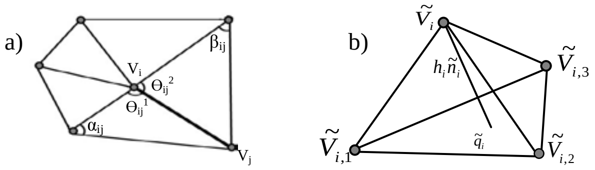


Figure 2. The angle used in cotangent weight and mean-value coordinate for edge (i , j). b) The one ring structure of the dual mesh in R^3 (R^6)

3.1.1 Discrete Laplacian Operators

A discrete Laplacian operator on M is determined by $L(v_i) = \sum_{j \in N(i)} w_{ij}(v_i - v_j)$ where $w_{ij} = \frac{1}{2}(\cot \alpha_{ij} + \cot \beta_{ij})$. As cotangent weight may be negative and problematic in very large angles case due to properties of the cotangent function be near π, thus convex weights that mimics cotangent weights are called mean-value coordinates [6] is determined as it follows:

$$w_{ij} = \frac{\tan(\frac{\theta_{ij}^1}{2}) + \tan(\frac{\theta_{ij}^2}{2})}{\|v_i - v_j\|} \tag{3}$$

The Laplacian operators are descibed above encode mesh details by using Laplacian coordinates. In order to better preserve mesh details in FS Laplacian deformation normal vectors of mesh vertices are taken into account.

3.1.2 Derivation of FS Laplacin Operator

As 1-ring structure of each vertex on dual mesh in R^6 is much simpler than the primal mesh, thus this kind of mesh can be easily used to compute Laplacian coordinate and they considered to derive FS Laplacian operator. The dual mesh $\hat{M} = (\hat{V}, \hat{E}, \hat{F})$ which is consisted of vertices positioned at centroids of triangle faces of the primal mesh will be constructed from given mesh M (V, E, F).
Suppose D be a full-rank matrix constructed from vertex face incidence matrix and is normalized, so that the sum of each row is equal to one. Thus the transformation from V to $\hat{V}$ is done as:

$$\hat{V} = DV \tag{4}$$

One of the important properties of the dual mesh is that the valence of every dual vertex is equal to three[7]. Thus, as shown in Figure 2.b, if the four vertices do not lie in one plane, then the 1-ring structure of dual mesh is essentially a tetrahedron. In Figure 2.b,$\tilde{V}_{i,1}, \tilde{V}_{i,2}, \tilde{V}_{i,3}$ are dual vertices in R^6, $\tilde{q}_i$ is perpendicular foot in the plane that contains the base triangle $\Delta \tilde{V}_{i,1}\tilde{V}_{i,2}\tilde{V}_{i,3}$, $\tilde{n}_i$ is the plane normal and $h_i \tilde{n}_i$ is the height vector of the 1-ring tetrahedron corresponding to the base triangle.
The extension of dual mesh from R^3 to R^6 after converting to dual mesh is necessary. Let $\hat{n}_i$ be the unit normal vector of face f_i of the primal mesh and w be a non-negative constant weight. There is a need to one to-one correspondence between vertices of the dual mesh and triangular faces of the primal mesh, for this aim each dual vertex $\hat{V}_i$ are mapped to the vertex $\tilde{V}_i = (\hat{V}_i, w\hat{n}_i)$ in R^6. As shown in Figure 2.b, a simple and unique representation for each dual vertex $\tilde{V}_i$ is provided with the 1-ring structure in dual mesh domain and in term of the plane which is defined by its 1-ring neighbors $\tilde{V}_{i,j}$, $j = \{1,2,3\}$:

$$\tilde{V}_i = \tilde{q}_i + h_i \tilde{n}_i = \sum_{j \in \{1,2,3\}} w_{i,j} \tilde{V}_{i,j} + h_i \tilde{n}_i \tag{5}$$

So that $w_{i,j}$ are the barycentric coordinate of $\tilde{q}_i$; they sum to one. With respect to equation (5), the Laplacian coordinate of a dual vertex is determined by:

$$\tilde{\delta}_i = h_i \tilde{n}_i = \sum_{j \in \{1,2,3\}} w_{i,j} (\tilde{V}_i - \tilde{V}_{i,j}) \tag{6}$$

So that the higher vector are used as the dual Laplacian coordinate of $\tilde{V}_i$.Since the normal of the base plane cannot be calculated by a cross product in R^6 we can still obtain the height vector of the tetrahedron. The base triangle $\Delta\tilde{V}_{i,1}\tilde{V}_{i,2}\tilde{V}_{i,3}$ determines a plane, expressed as:

$$\tilde{V}(u,v) = \tilde{V}_{i,1} + u(\tilde{V}_{i,2} - \tilde{V}_{i,1}) + v(\tilde{V}_{i,3} - \tilde{V}_{i,1}) \tag{7}$$

For determination of the height vector of tetrahedron, there is a need to find the perpendicular foot q_i of in the base plane, Thus

$$\begin{cases} (\tilde{V}_i - \tilde{q}_i).(\tilde{V}_{i,1} - \tilde{V}_{i,2}) = 0 \\ (\tilde{V}_i - \tilde{q}_i).(\tilde{V}_{i,1} - \tilde{V}_{i,3}) = 0 \end{cases} \tag{8}$$

From Equation (7), $\tilde{q}_i$ can be defined as $\tilde{q}_i = \tilde{V}_{i,1} + u_i(\tilde{V}_{i,2} - \tilde{V}_{i,1}) + V_i(\tilde{V}_{i,3} - \tilde{V}_{i,1})$ and the local coordinates (U_i , V_i) of $\tilde{q}_i$ in the base triangle is found by solving equation system(8). Finally, Laplacian coefficient $\tilde{w}_{i,j}$ can be determined from (U_i , V_i) by

$$w_{i,1} = 1 - U_i - V_i \quad , \; w_{i,2} = U_i \text{ and } \; w_{i,3} = V_i \tag{9}$$

Then FS Laplacian operator is obtained as:

$$L(\tilde{V}_i) = \sum_{j \in N(i)} w_{ij}(\tilde{V}_i - \tilde{V}_j) \tag{10}$$

In the matrix form $\tilde{\delta} = L\tilde{V}$, where L is Laplacian matrix

$$(L)_{ij} = \begin{cases} 1, & i = j \\ -w_{ij}, & j \in N(i) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

w_{ij} here and $w_{i,j}$ in Equation (9) have different form, but they have the same meaning. $w_{i,j}$ is used in 1-ring tetrahedron with j $\in$ {1,2,3} and w_{ij} is used in global mesh with j=1,2,…,n. As in this paper FS Laplacian operator is based on a FS metric, so FS Laplacian coordinate computation is affected with the unit normal vectors of the mesh vertices.

3.2 Tetrahedron Constraints

As, 1-rings structures can express local details of triangular meshes, so shapes of 1-ring structures should be preserved in order to preserve local details of meshes. Existing Laplacian deformation methods achieved this goal through preserving Laplacian coordinate of the mesh vertices. Since in a surface structure keeping volumetric sense is difficult and on the other hand, always the valence of each vertex in dual mesh is three, therefor as shown in Figure 2.b, 1-ring structure of the dual mesh can form a tetrahedron, which is called 1-ring tetrahedron. In order to construct a tetrahedron for each triangle, fourth vertex to each triangle is added [8]. The shape of the tetrahedron is preserved by minimizing both distortion of the base triangle and changing of the corresponding height of the 1-ring tetrahedron.

Let T'_{ti} be a tetrahedron, T_{ti} be the corresponding deformed one, $e_{j,k} = \hat{V}_{i,k} - \hat{V}_{i,j}$ be an edge in the base triangle Δ_i, $w^{T_{ti}}_{j,k}$ be the edge weight, R_i be the deformation matrix deforming T_{ti} to T'_{ti} and α gives the relative importance of the base triangle and the height vector of the 1-ring tetrahedron. Tetrahedron constraints are determined by following steps:

Step 1. Minimize the distortion of the base triangle. For this aim the energy function is defined by

$$E_{\Delta i} = \sum_{(j,k) \in (1,2)(2,3)(3,1)} w^{T_{ti}}_{j,k} \left\| e'_{j,k} - R_i e_{j,k} \right\|^2 \tag{12}$$

Step 2. Minimize the change of height h'_i by following energy function

$$E_{hi} = \left\| h'_i - R_i h_i \right\|^2 \tag{13}$$

So that $h_i = L(\hat{V}_i)$ are exactly FS Laplacian coordinates, which are given in equation (10). In fact, the height constraint here is a Laplacian coordinate constraints. A weighted combination of the above is minimized to preserve the shape of T_{ti} as follow:

$$E_{T_{ti}} = E_{hi} + \alpha E_{\Delta i} \tag{14}$$

The carefully determined weight factor α in equation (14) can present the shape of T_{ti} can be preserve the shape of as much as possible.

3.3 Deformation with Tetrahedron Constraints

Let $E_{T_{ti}}$ be tetrahedron constraint which is calculated in equation (14), V_i^c, i=1,2,…, m are the constraints vertices, M' be deformed mesh and β is set to a relatively large value in order to move handle vertices conveniently. To deform a mesh model, the user first deforms the handle vertices. Fixed vertices and handle vertices serve as position constraints. Moreover, it has been observed that the solution behaves better when position constraints are satisfied in least square scene rather than the mesh are deformed by solving quadratic minimization problem

$$\min_{v_i'}(E_{T_{ti}} + \beta \sum_{i=1}^{m} \left\| V_i^c - c_i \right\|^2) \tag{15}$$

So that in equation (15) positional constraints of fixed and handle vertices are soft constraints. From equation (12)-(15) the energy function will be minimized and defined as follows:

$$E(M') = \sum_{i=1}^{n} (E_{h_i} + \alpha E_{\Delta i}) + \beta \sum_{i=1}^{m} \left\| V_i^c - c_i \right\|^2 \tag{16}$$

3.3.1 Determining the Weight

As h_i are the Laplacian coordinates, the Laplacian operator is used to determine w_{ij},see equation (9). On the other hand by using the FS Laplacian operator, we can obtain excellent feature preservation $w_{j,k}^{Tti}$ which is the edge weight of the base triangle is determined by averaging weight of two adjacent edges $w_{i,j}$ *and* $w_{j,k}$.

3.3.2 Propagation of Local Transforms R_i

Let F be a harmonic function on the primal mesh (f_i for vertex V_i), L be FS Laplacian and D be the primal to dual mesh. Given the local deformation R_h of the handle mesh, in natural way the interpolation over whole mesh should be done in order to give the overall mesh deformation. For discrete scalar harmonic fields on the primal mesh, the value 1, is assigned as a boundary condition at handle vertices and 0 is given for fixed vertices. Harmonic fields on the primal mesh can be determined by solving in a least-squares sense as follow:

$$LDF = 0, V_{fixed} = 0 \quad \text{and} \quad V_{handel} = 1 \tag{17}$$

So the harmonic fields on the dual mesh can be obtained by $\hat{F} = DF$.The deformation R_i can be computed form the local deformation by using

$$R_i = f_i R_h \tag{18}$$

3.3.3 Energy Minimization

Let R be the global rotation which is applied to the entire mesh, the Ms be edge matrices, Es be edges of the base triangle of the tetrahedron, $c_{1:m}$ be positional condition for handle and fixed vertices, L' be the whole coefficient matrix and b is the whole right hand side vector. The energy function can be rewritten in matrix form as:

$$\begin{bmatrix} L \\ \alpha M_{e12} \\ \alpha M_{e23} \\ \alpha M_{e31} \\ \beta I_{m\times m} \end{bmatrix} \hat{V} = \begin{bmatrix} R\hat{\delta} \\ \alpha RE_{12} \\ \alpha RE_{23} \\ \alpha RE_{31} \\ \beta c_{1:m} \end{bmatrix} \tag{19}$$

With equation(4) and (19), the energy function can be rewritten by:

$$L'DV = b \tag{20}$$

This is linear system AV=b. The Cholesky factorization is applied to solve the system in order to associate normal equation:

$$A^T AV = A^T b \tag{21}$$

First a sparse Cholesky factorization A^TA is computed. Then, the deformed vertices are obtained by back substitution.

3.4 Algorithm Summery

The main steps of the proposed FS deformation for triangular mesh include: First, obtain the FS Laplacian matrix L according to section 3.1.2 and edge matrices M_{e12}, M_{e23}, $M_{e31.}$ Then, assuming handle's rotation and positional condition are R_h and $c_{1:m}$, select fixed vertices and handle vertices and deform the handle. Next, use harmonic guidance method according to section 3.3.2 to propagate the rotation R_h to the whole mesh. Finally, apply equation (20) for reconstructing mesh models.

4. CONCLUSION

In this paper, an approach to evaluate footwear fit within a shoe last data base is proposed. To fulfill the task, an algorithm for cutting the 3D models to several sections to find the footwear within a shoe last data base is introduced. Then the area of the contour which is established from 3D costumer foot is calculated and searched for equal section in shoe last data base to find the similar shoe last's contour. If that shoe last's contour has the same shape and the same area that one is chosen as the best fit. Otherwise shoe last's contour is deformed with respect to customer foot's contour. This new approach should clearly help to improve the user's comfort and it could be the starting point for new types of mass customization approach in footwear design.

REFERENCES

[1] B. Ture Savadkoohi, R. De Amicis, "A CAD System for Evaluating Footwear Fit", Multimedia, Computer Graphics and Broadcasting, Vol.60, pp.1-7, 2009.

[2] B.Ture Savadkoohi, R. De Amicis, "Mass Customization Strategy In Footwear Industry", International Confrence On Advances In Information Processing And Comunication Technology(IPCT), pp. 12-16, ROME, Italy, 18-19 April, 2015.

[3] J. Gain and D. Bechann, "A Survey Of Spatial Deformation From User-Centred Prespective", Journal of ACM Transactions On Graphics (TOG), Vol.27, No.4, pp.1-21, 2008.

[4] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez, B. Levy, "Polygon Mesh Processing", AK Peyers, ISBN 978-1-56881-426-1, 2010.

[5] J. He, C. Zhang, Y. We and W. Li, "Feature Sensitive Deformation For Triangular Mesh Models", Journal of Computer Animation and Virtual Worlds, Vol.22, No.1, pp. 15-25, 2010.

[6] MS. Floater, "Mean value coordinates", Journal of Computer Aided Geometrical Designing, Vol.20, No.1,19-27, 2003.

[7] G. Taubin, "Dual Mesh Resampling", In PG'01, Proceedings of the 9th Pacific Confrence On Computer Graphic and Applications, pp.180, 2001.

[8] H.B. Yan, S. Hu, R.R.Martin, Y.-L.Yang, "Shape Deformation Using A Skeleton To Drive Simplex Transformation", Jornal of IEEE Transaction on Visulaization And Computer Graphic, Vol.14, No.3, pp.693-706, 2008.

Reflection Papers

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

GAZE-EMBODIED AND GAZE-OFFLOADED COGNITION: IMPLICATIONS FOR GAZE-CONTROLLED INTERFACE DESIGN

Nadiya Slobodenyuk
American University of Beirut, Department of Psychology, P.O.Box 11-0236, Riad El-Solh / Beirut 1107 2020, Lebanon

ABSTRACT

The paper presents key findings on the relationship between eye movements and cognition followed by discussion of the implications of restriction of natural eye movements on mental operations and the implications of guided generation of task-decoupled eye movements. Specific aspects of gaze-controlled interface design are then addressed from perspective of eye movement suppression and guided generation. The analysis suggests that the effects of altering natural task-relevant eye movements in gaze controlled interfaces are not negligible and that the user concerns should be extended beyond the typical usability issues to include information processing effects.

KEYWORDS

Gaze-controlled interfaces, cognitively-grounded design, embodied cognition

1. GAZE AND COGNITIVE EMBODIMENT

The notion of embodied cognition refers to the strong link between the mind and the acting body and presumes that the interaction of the body with the environment itself to some extent determines and gives the rise to cognition. Various theories of embodied cognition emphasize different aspects of embodiment, including the role of bodily states in cognition, and sensorimotor simulation that accompanies conceptual processing. The evidence of the close link between the motor system and cognition comes from a vast variety of mental processes, and a particularly strong argument of cognitive embodiment is being built on findings in the area of linguistic processing. For example, Brouillet et al. (2010) have shown that “yes” verbal response is associated with faster pushing of the lever while “no” response with faster pulling of the lever. Such implicit motor component could be explained by automatic motor planning for reaching and withdrawal associated with verbal responses. Other prominent linguistic examples include embodied processing of action verbs (Boulenger et al. 2006), embodied sentence comprehension (Glenberg & Kaschak 2002), and even comprehension of abstract concepts (Lakoff & Johnson 1999). Similar motor biases have been shown in numerical processing (Fischer 2012), music cognition (e.g., Leman 2007), moral reasoning (e.g., Schnall et al. 2008) and many other processes (see Shapiro 2014, for overview). The embodied information processing is also supported by neuroimaging studies (Chao & Martin 2000) and studies on cognition in cases of compromised motor system (Neininger & Pulvermüller 2003; Boulenger et al. 2008).

Most evidence of cognitive embodiment comes from behavioral paradigms that employ simple motor response (e.g., key press), therefore, pointing to the link between bodily action and cognition. Less is known about cognition embodied in eye movements. What is, however, clear is that eye movements in addition to being a literal embodiment of visual attention (Shapiro 2014, p. 311) are also closely linked to other cognitive functions through the motor system. Such link has been primarily shown in the studies on eye movement patterns characteristic to mental operations. Lorens and Darrow (1962) have shown increase in eye movements during mental calculation. Nitschke et al. (2012) recorded eye movements while their participants were solving The Tower of London (TOL) task. TOL is a disc-transfer task in which the manipulation of the tower configuration allows altering the demands on establishing mental representation and internalization of the problem. It also allows altering the demands on planning processes. Nitschke et al.’s results showed that eye movement parameters (i.e. number of gaze alteration and fixation duration) were

ISBN: 978-989-8533-38-8 © 2015

determined by differential task demands on the processes of internalization and planning. Others have shown that mental workload itself affects eye movements in a restricting manner (May et al. 1990).

Eye movements not only embody mental processes but also allow decreasing the demands on cognitive resources by offloading some mental operations onto the environment. For example, Ballard et al. (1995) have notably shown that while participants could rely on memory to perform on a block-copying task that required replication of the configuration of the colored blocks shown in the model, they actually serialized the task with eye movements thereby decreasing the use of short-term memory.

2. CONSIDERATIONS FOR GAZE-CONTROLLED INTERFACE DESIGN

Despite the robust evidence of the strong link between cognition and eye movements, such consideration has not yet found its way into the field of gaze-controlled interface design. This gap in the design considerations could be justified if we assume that the connection between mental operations and natural eye movements that accompany theses operations is negligible. Under such premise, we would also need to assert that eye movements produced for the purpose of interface control do not interfere with cognition embodied in natural eye movements and that gaze restrictions imposed by the way gaze is employed in the interface (e.g., in case of gaze gesture) do not significantly alter task-relevant cognition. Some insight on whether this is true can be drawn from evidence of the effect of saccadic suppression on cognition. Wallentin et al. (2011) have shown that eye movement suppression interferes with working memory and suggested that the effect is due to interference between eye movements and the process of manipulation of mental representations. Postle et al. (2006) have shown interference effects in visuospatial processing and proposed that the interference is not due to eye movements per se but rather due to eye movement control. This is a particularly important finding considering that gaze-controlled interfaces inherently lead to alterations in effortless and consciously inaccessible eye movement control towards a more controlled and deliberate process.

If suppression of naturally occurring eye movements during mental operations interferes with information processing, guided generation of eye movements can both hinder and facilitate cognition. Lawrence et al. (2004) using primary and secondary spatial working memory tasks that required shifts of attention, shifts of eye movements, or no shifts, have shown that some interference can be caused by attention shifts that accompany eye movements, however, they have also noted a separate type of interference that is caused by eye movements alone. Another line of research has shown that saccadic movements unrelated to the task can enhance performance. One of the well supported effects here is horizontal saccade-induced retrieval enhancement. This effect refers to the finding that generation of horizontal saccadic eye movements after memory encoding enhances memory performance at the retrieval stage and might be due to increased functional connectivity between the two hemispheres (Nieuwenhuis et al. 2013).

The discussed examples come from research on the relationship between cognition and eye movements outside the context of gaze-controlled interface design. In the domain of gaze-controlled interfaces, user experience is typically addressed in a relatively general manner and the experience of ease of control is considered to be representative of most user concerns. Considerations of the information processing consequences are yet scarce, however, even little knowledge on cognitive performance in gaze-controlled interfaces there is points to the fact that employing gaze for interface control could have profound effects on task-related cognition. For example, Bednarik et al. (2009, p. 5) investigated differences in problem solving using gaze selection and mouse selection. Their study has shown that participants were able to adjust to the limitations imposed by the gaze-controlled interface design. In particular, participants developed efficient strategies to avoid undesirable selections; however, these strategies necessary for efficient interface control also resulted in some participants feeling that they could not “think freely”.

Taking into account the discussed findings, several implications for gaze-controlled interface design can be noted. First, research confirms that gaze restrictions could hinder some working memory processes and increase the cognitive load by limiting the natural gaze-offloading strategy. These effects are particularly important considering that gaze restrictions have been used as a strategy to avoid the Midas Touch problem and solve the problem of undesirable selections in general. A particular example here would be dwell activation. A widely used continuous dwell ensures that the activation is intentional; however, it also interrupts the natural eye movements through spatial restriction. Taking into account the non-discrete, cascade-like nature of information processing, it is likely that dwell-based eye movement restriction necessary for command generation interferes with some online cognitive processes. Second, generation of multiple task-decoupled gaze shifts can significantly alter spatial processing and encoding of information. A particular attention to this issue should be given during gaze gesture design. Gaze gesture is becoming an excellent alternative to dwell activation as it allows immediate command generation. Simple one-segment and complex two- and three-segment gaze gestures have been tested for suitability for a range of commands, including moving and resizing the objects (Heikkilä 2013) and complex game control (Istance et al. 2010). The cognitively aware approach proposed here would include testing which gaze-gestures lead to less cognitive interference and, therefore, should be adopted. By analogy to stimulus-response compatibility principles we are currently using for hand-response design, identification of stimulus-gaze-response compatibility principles could be the first step towards gaze gesture that does not interfere with task-related information processing. Finally, both restriction and generation of eye movements for the purpose of interface control could influence the ratio between the processes of planning, internalization and externalization in problem solving. Whether the change in the ratio between these processes places serious demands on cognitive resources is still unclear, however, in certain types of gaze-controlled applications, the matter should most certainly be addressed.

3. CONCLUSION

Empirical evidence of the relationship between eye movements and cognition points to the fact that eye movement suppression and task-decoupled eye movement generation can have profound effects on the cognizer. So far these effects were addressed systematically in cognitive psychology but did not make their way into the applied field of gaze-controlled application design. The evidence, nevertheless, strongly suggests that the cognitive effects are not negligible and should be studied along other user concerns. Such extension of the user considerations to include cognitive effects in the field of gaze-controlled interface design will allow creating interfaces that ensure that eye movements produced for the purpose of interface control do not interfere with cognition embodied in natural eye movements. Additionally, evidence of saccadic facilitation of cognition indicates that information processing considerations will not only allow avoiding undesirable effects, but will also allow approaching gaze-controlled interaction from perspective of cognitive benefits beyond the ones offered by other types of interfaces.

REFERENCES

Ballard D. et al, 1995. Memory representations in natural tasks. *Cognitive Neuroscience*, Vol. 7, pp 66–80.

Bednarik R. et al, 2009. Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience. *Journal of Eye Movement Research*, Vol. 3. No. 1, pp 1–10.

Boulenger V. et al, 2006. Cross-talk between language processes and overt motor behavior in the first 200 msec of processing. *Journal of Cognitive Neuroscience*, Vol. 18, No. 10, pp 1607–1615.

Boulenger V. et al, 2008. Word processing in Parkinson's Disease is impaired for action verbs but not for concrete nouns. *Neuropsychologia*, Vol. 46, pp 743–756.

Brouillet T. et al, 2010. The embodied cognition theory and the motor component of "yes" and "no" verbal responses. *Acta Psychologica*, Vol. 134, pp 310–317.

Chao LL. & Martin A., 2000. Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, Vol. 12, pp 478−484.

Fischer MH., 2012. A hierarchical view of grounded, embodied, and situated numerical cognition. *Cognitive Processing*, Vol. 13, pp S161–S164.

Glenberg AM. & Kaschak MP., 2002. Grounding language in action. *Psychonomic Bulletin & Review*, Vol. 9, pp 558–565.

Heikkilä H., 2013. Tools for a gaze-controlled drawing application - Comparing gaze gestures against dwell buttons. *Lecture Notes in Computer Science*, Vol. 8118, pp 187–201.

Istance H. et al, 2010. Designing gaze gestures for gaming: an investigation of performance. *Proceedings of the 2010 Symposium on Eye-tracking Research & Application*. Austin, TX, USA, pp 323–330.

Lakoff G. & Johnson M., 1999. *Philosophy in the flesh: The embodied mind and its challenge to western thought*. Basic Books, New York, USA.

Lawrence B. et al, 2004. Interference with spatial working memory: An eye movement is more than a shift of attention. *Psychonomic Bulletin & Review*, Vol. 11, pp 488–494.

Leman M., 2007. *Embodied Music Cognition and Mediation Technology*. MIT Press, Cambridge, USA.

Lorens SA. & Darrow CW., 1962. Eye movements, EEG, GSR and EKG during mental multiplication. *Electroencephalography and Clinical Neurophysiology*, Vol. 14, No. 5, pp 739–746.

May JG. et al, 1990. Eye movement indices of mental workload. Acta Psychologica, Vol. 75, pp 75–89.

Neininger B. & Pulvermüller F., 2003. Word-category specific deficits after lesions in the right hemisphere. *Neuropsychologia*, Vol. 41, pp 53–70.

Nieuwenhuis S. et al, 2013. Bilateral saccadic eye movements and tactile stimulation, but not auditory stimulation, enhance memory retrieval. *Brain and Cognition*, Vol. 81, pp 52–56.

Nitschke K. et al, 2012. Dissociable stages of problem solving (I): Temporal characteristics revealed by eye-movement analyses. *Brain and Cognition*, Vol. 80, pp 160–169.

Postle BR. et al, 2006. The selective disruption of spatial working memory by eye movements. *Quarterly Journal of Experimental Psychology*, Vol. 59, No. 1, pp 100–120.

Schnall S. et al, 2008. Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, Vol. 34, pp 1096-1109.

Shapiro L., 2014. *The Routledge handbook of embodied cognition*. Routledge, UK.

Wallentin M. et al, 2011. Eye movement suppression interferes with construction of object-centered spatial reference frames in working memory. *Brain and Cognition*, Vol. 77, pp 432–437.

VISUALISATIONS AND THEIR EFFECT ON COGNITIVE BIASES IN THE CONTEXT OF CRIMINAL INTELLIGENCE ANALYSIS

Eva-Catherine Hillemann, Alexander Nussbaumer and Dietrich Albert
Knowledge Technologies Institute, Graz University of Technology, Inffeldgasse 13/5, A-8010 Graz, Austria

ABSTRACT

Criminal intelligence analysts of today are confronted with a veritable explosion in data volume. They are constantly required to make sense of these data and consequently to make appropriate decisions. To effectively manage the complexity of information and to avoid to be overwhelmed by it, humans unconsciously apply heuristics. These heuristics, although generally useful, can lead to systematic errors in judgment, so-called cognitive biases. Research in the area of information visualisation has shown a positive effect of visualisations on the sense-making process by supporting humans in their understanding and interpreting of large data sets. This paper gives an overview of the current research in the following two areas (1) cognitive biases with its mitigation strategies and (2) information visualisation. This paper also demonstrates that more research is necessary to link these both research areas together by investigating the effect of visualisations on cognitive biases during the sense-making process.

KEYWORDS

Sense-making, cognitive bias, confirmation bias, bias mitigation, information visualisation

1. INTRODUCTION

In our technology-driven society, managing the flood of available data has become tremendously challenging. A criminal intelligence analyst has to identify relevant information from very large datasets and has to piece all relevant information together so that it is possible to draw a sensible, reasonable, justifiable, and defensible conclusion. Sense-making in this context means that an analyst has to actively construct a meaningful and functional representation of some aspects of the whole picture. This allows him for getting new understanding and new insight. Sometimes this process is influenced by prior beliefs or experiences that may lead to wrong decisions. To understand such failures in human sense-making, a lot of research especially in the field of cognitive psychology has been done. So-called cognitive biases have been proposed as possible explanation for such pitfalls in judgment and decision-making (Kahneman and Tversky, 1972).

Especially in the field of criminal intelligence analysis, it is essential to make a right decision as decision making failures can have severe consequences such as wrongful convictions. Thus, main challenge that arises is to present data in such a way that sense-making is supported and failures are avoided. Providing visualisations is an appropriate method as it combines technology-enhanced analysis techniques with cognitive-analytical abilities of human being. This is precisely the point where the European project VALCRI (www.valcri.org) comes into play. Addressing the challenges of todays' law enforcement agencies, the main purpose of this project is create a system that will support human reasoning and sense-making by providing appropriate data analytic tools following the principles of visual analytics.

From a cognitive bias research point of view, it is of particular importance to investigate the influence and effect of visualisations on cognitive biases in the sense-making process. The following research question are of main interest: (1) Which effect do visualisations have on cognitive biases during the sense-making process? Which cognitive biases are activated, which cognitive biases are reduced?, and (2) How and in which way have visualisations to be designed and applied in order to mitigate/reduce the occurrence of cognitive biases?

This paper aims at giving an overview of the current research in the fields of cognitive biases and data visualisation that serves the theoretical basis for being able to answer these research questions. It is structured as follows: Section 2 and Section 3 give an overview of the relevant state of the art and theoretical background in the field of visual representations on the one hand, and in the field of cognitive biases with a focus on the confirmation bias and its mitigation on the other hand. Finally, Section 4 presents concluding remarks and future work.

2. COGNITIVE BIASES

To effectively manage information collection and processing as well as to simultaneously avoid to be overwhelmed by too much information, humans unconsciously apply "heuristics". Such "rules of thumb" or short cuts are primarily used to simplify a cognitive task in order to allow decision making when a fast decision is required and time and resources are limited (Klein, 1999). Thus, they do not give optimal solutions, just "good enough" solutions. Applying heuristics allows humans to save efforts and time, but sometimes at the cost of accuracy (Payne, Bettman, & Johnson, 1993; Shah & Oppenheimer, 2008). Such resulting "systematic errors" in judgment and decision making are referred to as cognitive biases. Tversky and Kahnemann (1974) introduce the term "cognitive bias" in their heuristics-and-bias program as "a pattern of deviation in judgment that occurs in particular situations, leading to perceptual distortion, inaccurate judgment, illogical interpretation, or what is broadly called irrationality".

In the context of criminal intelligence analysis, the most-well known information processing bias is the confirmation bias. This bias addresses the tendency to search for or interpret information in a way that confirms one's preconceptions or hypotheses. Humans tend to seek confirmatory evidence while not searching for disconfirming information. Furthermore, uncertainty in gathered, collated, and analysed information can lead to biased interpretation of this information. Research on confirmatory hypothesis testing indicates that preliminary expectations of a person can have an impact on visual as well as auditory perception (Lange, Thomas, Dana, & Dawis 2011). Confirmation bias in this context is a natural and automatic feature operating without conscious awareness (Findley & Scott, 2006) and without a person's self -interest (Nickerson, 1998).

In the last years, increased attention has been given to the development of interventions to reduce information-processing errors. Wilson and Brekke (1994) suggested that being aware might facilitate the judgment adjustment of decision makers. Such awareness can be imparted by a training program where different methods such as feedback, coaching, offering warnings about the possibility of bias, and describing the direction of a bias can be used to improve the decision making process (Fischoff, 1982). Focusing on the confirmation bias and its mitigation, a more or less effective strategy is to ask people to "consider the opposite" (Anderson & Sechler, 1986; Lord, Lepper, & Preston, 1984) or to generate any alternative (Hirt, Kardes, & Markman, 2004; Hirt & Markman, 1995). Generatingcounterarguments to why their belief may be wrong (Sanna & Schwarz, 2004) can have a positive influence on reducing the occurrence of the confirmation bias. Hirt and Markman (1995) argues in their work, that the more plausible such an alternative explanation is for people, the more effective is the value of debiasing. The most prominent example for the "considering the opposite" strategy is the eight-step "Analysis of Competing Hypotheses" (ACH) technique for intelligence analysts developed by Heuer (1999). The main idea of ACH is to reduce cognitive biases by encouraging analysts to consider all reasonable hypotheses when evaluating a set of evidences (Heuer, 1999).

Unfortunately, such interventions or mitigation strategies have often reached no satisfying result. So there is a need for additional as well as effective mitigation strategies that help the analyst to overcome or at least minimise their biases.

3. INFORMATION VISUALISATION

When people have to make sense of or interpret a large amount of different data, the main question that arises is how and in which way all these data can be presented in order to support the underlying cognitive processes. Presenting data through visualisations, "graphical representations of data and concepts" (Ware, 2012), allows for a better and easier understanding and interpreting of a very large data set in a very quick way (Edelson, Pea, & Gomez, 1996; Wood, Wright, & Brodlie, 1997). Holistic visual representations, so-called information visualisations - help people to generate new insights and thus improve their problem-solving success (Larkin & Simon, 1987; Dull & Tegarden, 1999). Card, Mackinlay, and Shneiderman (1999) outlines the benefits of the application of visualisation techniques as follows: i) increasing memory and processing resources available to the user, ii) reducing the search for information, iii) enhancing the detection of patterns, iv) enabling perceptual interference patterns, v) using perceptual attention mechanisms for monitoring, and vi) encoding information in a manipulable medium. Spence (2011) states that information visualisation can be described as a cycle process whereby the visualisation leads on to a mental model, a so-called cognitive map, that supports the sense-making process. The creation and interpretation of such a cognitive map is based on the following four activities: i) browsing (i.e. register the content), ii) modelling (i.e. browsing outcome is integrated to form an internal model), iii) interpretation (i.e. deciding how and whether further browsing should continue), and iv) formulating browsing strategies.

When having a look at the current research literature in the field of visualisations, it becomes clear, that most of the research topics focus on human-computer interaction. This includes both psychological research on the perception of visual information and research on how to best design the user interface (Bederson & Shneidermann, 2002). For effectively using visualisations, an appropriate design, that follows design guidelines for instance based on the assumptions of Gestalt psychology, is undoubtedly important. However, little research has been done in investigating the influence and effectiveness of visualisation on cognitive processes including sense-making and the occurrence of cognitive biases. In the context of intelligence analysis, one of the few studies is the one of Billman et al. (2006). They explore the influence of visualisations on the occurrence of cognitive pitfalls such as the confirmation bias. The most interesting result is , that the ability to explore new perspective through using different visualisation types may lead analysts to develop more diverse and numerous hypotheses. Cook and Smallman (2007) found, that graphical layout of evidence has a negative impact on the occurrence of the confirmation bias.

4. CONCLUSION AND NEXT STEPS

In criminal intelligence analysis, it is essential to make right decisions and to appropriately make sense of a full range of data. Errors can have severe consequences, thus, there is a need to reduce and mitigate the probability of their occurrence. Many studies in the research of cognitive biases and their mitigation show that training alone has not much effect on the mitigation of biases.Additionally, in many cases awareness of the bias does not produce a more accurate perception. Research on visualisation has shown that multiple views on data might suppress biases before they occur. This change of visualisation might induce a change in the thinking process. In the context of the project VALCRI a main focus is on bringing together both research areas: cognitive bias and information visualisation.. Future research will examine the impact of information visualisation on the occurrence of cognitive biases, especially the confirmation bias. Thus, a psychologically based empirical experimental study is planned with a specific focus on investigating whether presenting information through different visualisation types has an impact on both the sense-making process of analysts in the criminal intelligence context, in general, and the occurrence of the confirmation bias, in particular.

ACKNOWLEDGEMENT

The research leading to these results in project VALCRI has received funding from the European Union 7th Framework Programme (FP7/2007-2013) under grant agreement no FP7-IP-608142.

REFERENCES

Anderson, C.A., Sechler, E.S. (1986). Effects of explanation and counterexplanation on the development and use of social theory. *Journal of Personality and Social Psychology*, Vol.50, pp. 24-24.

Bederson, B.B., & Shneiderman, B. (2002). *The Craft of Information Visualization: Readings and Reflections.* Morgan Kaufmann.

Billman, D., Convertino, G., Shrager, J., Pirolli, P., & Massar, J. (2006). Collaborative intelligence analysis with CACHE and its effects on information gathering and cognitive bias. *Human Computer Interaction Consortium Workshop*

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. New York, NY: Morgan-Kaufmann.

Cook, M. B., & Smallman, H. S. (2007). Visual evidence landscapes: Reducing bias in collaborative intelligence analysis. *Proceedings of Human Factors and Ergonomics*, Vol. 51, pp. 303–307.

Dull, R. B., & Tegarden, D.P. (1999). A comparison of three visual representations of complex multidimensional accounting information. *Journal of Information Systems*. Vol. 13, No. 2 (Fall), pp. 117-131

Edelson, D., Pea, R., Gomez, L. (1996). Constructive in the collaboratory. In B.G. Wilson (Ed.). *Constructivist learning environments: Case studies in instructional design*. Englewood Cliffs, NJ: Educational Technology Publications.

Findley, K. A., & Scott, M. S. (2006). The multiple dimensions of tunnel vision in criminal cases. *Wisconsin Law Review*, Vol. 2, pp. 291–297. Retrieved from http://hosted.law.wisc.edu/lawreview/issues/2006 –2/findley-scott.pdf

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgement under uncertainty: Heuristics and biases, 422-444. Cambridge, England: Cambridge University Press.

Heuer, R. J. Jr.. (1999). Psychology of intelligence analysis. Washington, DC: Central Intelligence Agency Center for the Study of Intelligence. Retrieved from https://www.cia.gov/library/center-for-the-studyofintelligence/csi publications/books-and-monographs/psychology-ofintelligence-analysis/index.html

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mindset through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology,* Vol. 40, pp. 374-383.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, Vol. 69, pp.1069-1086.

Hoover, J. D., & Whitehead, C. (1975). An Experiential-Cognitive Methodology in the First Course in Management: Some Preliminary Results. In Richard H. Buskirk (ed.). Simulation Games and Experiential Learning in Action, pp. 25-30.

Klein, G. (1999): *Sources of Power: How People Make Decisions*, MIT Press, Cambridge, Massachusetts.

Lam, C.-M. (2007). Is Popper's Falsificationist Heuristic a Helpful Resource for Developing Critical Thinking?, *Educational Philosophy and Theroy*, Vol. 39 (4), pp. 432-448.

Lange, N. D., Thomas, R. P., Dana, J., & Dawes, R. M. (2011) Contextual biases in the interpretation of auditory evidence. *Law and Human Behaviour, Vol. 35,* pp.178–187.

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, Vol. 11, pp. 65–100.

Lord, C.G., Lepper, M.R., & Preston, E. (1984). Considering the opposite: a corrective strategy for social judgment. *JPers Soc Psychol*, Vol. 47, pp. 1231-1243.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, Vol. 2, pp. 175–220

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The adaptive decision maker. New York: Cambridge University Press.

Sanna, L. J., & Schwarz, N. (2004). Integrating temporal biases: The interplay of focal thoughts and accessibility experiences. *Psychological Science,* Vol. 15, pp. 474-481

Shah, A. K., & Oppenheimer, D. M. (2007). Easy does it: The role of fluency in cue weighting. Manuscript under review.

Speier, C., & Morris, M.G. (2003). The influence of query interface design on decision-making performance. *MIS Quarterly.* Vol. 27, No. 3, pp. 397-423

Spence, R. (2001). *Information Visualization.* Addison-Wesley.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, Vol. 185, pp. 1124–1131.

Wood, J., Wright, H., & Brodlie, K. (1997). Collaborative visualization. *Proceedings of Vis1997*. IEEE Press.

Ware, C. (2012). *Information visualization: perception for design*. Elsevier.

Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science,* Vol. 11, pp. 65–100.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin,* Vol. 116, pp. 117–142.

Posters

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

BRAIN-COMPUTER INTERFACE APPROACH TO COMPUTER-AIDED DESIGN: ROTATE AND ZOOM IN/OUT IN 3DS MAX VIA IMAGINATION

Yu-Chun Huang[1], Kuan-Lin Chen[2], Ming-Ying Wu[1], Ying-Wei Tu[1]
and Scottie Chih-Chieh Huang[3]
[1]*College of Design, Tatung University*
[2]*Dept. of Electrical Engineering, National Taiwan University*
[3]*Department of Industrial Design, Chung Hua University*

ABSTRACT

To increase the efficiency of 3D modeling, the sophisticate CAD users must memorize different hotkey combinations for different softwares. Hence, the challenge of the research is to find out how to create a more intuitive and natural “rotate” and “zoom in/ out” command in 3D CAD modeling. The research creates a “BCI embedded CAD” system prototype via EPOC+ with the connection between 3D studio Max and brainwaves. Through the system, the user can easily control the 3D viewport (zoom in /out) in 3D CAD environment through “thinking the commends”. In the future study, the “BCI embedded CAD” can be modified as a cross-platform 3D CAD interface that enables users to use imagination to control 3D modeling among different softwares (Maya, 3D studio Max, and Rihno) rather than traditional text-based commands or graphical icons.

KEYWORDS

Human-Computer Interaction Interface; Brain-Computer Interface; Computer-Aided Design.

1. INTRODUCTION

Sculpturing a physical model in real world is easy and common for a designer. However, in the 20st century, with the development of “Computer-Aided Design/ Computer-Aided Manufactured (CAD/CAM)” (Lichten, 1984), modeling has increased efficiently, but the other problem came along to untrained users. The users need to learn how to use the 3D software via complicated and inconvenient interfaces, especially for the novices. Thus, researchers focus on the 3D input problem for many years (Aish, 1979). Aish (1979) aiming that 3D input systems should be able to create and modify 3D geometry intuitively in order to interpret and evaluate the spatial qualities of a design directly. However, in the 1980s, the CAD interface was limited to text-based commands; in the 1990s, the windows, icons, menu and pointers-based (WIMP) and text hybrid dominated CAD system. We can see that in most 3D modeling systems, text-based command system and Graphic-User Interfaces (GUIs) are still the mainstream. The keyboard and mice are essential for users to type in and select commands, such as the first CAD software—“Sketchpad” (1963, Sultherland), and recently, CAD software—“Maya”, “3D studio Max” and “Rihno”.

2. PROBLEM AND OBJECTIVE

Regarding to the traditional HCI interfaces (keyboard, mice, gesture and pen), users must spend time learning how to manipulate 3D modeling among different softwares (such as Maya, 3D studio Max and Rihno). For instance, the commands—“viewport rotating and zoom in/ zoom out the object” that the users use frequently in 3D modeling world, users have to memorize the multiple hotkey (keyboard + mice) to reach the goal (rotating object) or use the graphic icon to finish the action (Figure 1-left). Moreover, sometimes it is harder if you cross different 3D modeling softwares (Maya and 3D studio Max). The rotation function hotkeys are totally different even if there are designed in the same company (Autodesk). To increase the efficiency of 3D modeling, the CAD users must memorize different hotkey combinations for different softwares. Hence, the challenge of the research is to find out how to create a more intuitive and natural "rotate" and "zoom in/ out" command in 3D CAD modeling. Also, this alternative "rotate" and "zoom in/ out" input can be widely applied to different 3D CAD softwares. Brain-computer interface (BCI) has been applied to the various fields: BCI embedded robot arm control (Chapin et al., 1999); BCI game (Krepki, 2007) BCI embedded smart space design (Huang, 2011).

To improve the unintuitive and unfriendly interface of 3D CAD software (eg. Maya and 3D studio Max), this research combines BCI into CAD system to create a better interface in 3D CAD manipulation. By monitoring user's brainwaves generated when intending to perform different commands (see figure 1-right-B), the user is able to easily control the 3D viewport "zoom in/out" command through imagining zoom in/out instead of relying on the traditional input commands (keyboard + mice or graphical icons).

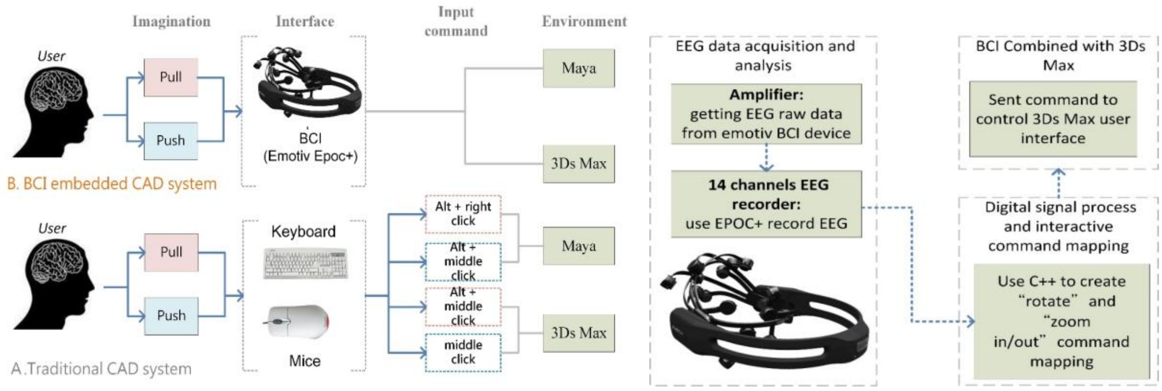


Figure 1. Left) The comparison of BCI CAD system and traditional CAD system; right) BCI CAD system framework.

3. METHOLOGY AND STEPS

In order to implement the "BCI embedded CAD" system, the methodology of the research is divided into three steps: the first step: the BCI training process via EPOC+ for new users; the second step, BCI embedded CAD system implementation; the third step, scenario demonstration and discussion:

The First Step, BCI Training Process via EPOC+ for New Users: In the past BCI devices, a new user's brainwaves could be recognized for high accuracy if the user went through a long time training process to achieve the goal. In order to ensure that the EPOC+ can be widely for every users, the first step of this research is to find out how to create a short and effective training process for new users. After the training process, the user can achieve 90% accurate control the virtual object in 3D environment.

The Second Step, System Implementation: This step divided into three parts, EEG data acquisition and analysis; digital signal process and interactive command mapping; BCI combined with 3Ds Max by using C++ and maxscript (see figure 1-right).

The Third Step, Scenario Demonstration: In order to demonstrate the system prototype the subjects are asked to build a box and try to use the imagination command to zoom in/ out the box in 3Ds Max.

4. BCI TRAINING PROCESS VIA EPOC+ FOR NEW USERS

The first, EPOC+ Installation: in order to acquire the accurate signal, the subject would follow the steps to install the EPOC+ system. First the user has to wear the EPOC+ to his head. Second the user opens the EPOC+ software to check the contact quality of each sensor. The sensor electrode figure appeared the quality of the contact quality from green, yellow, orange, red and black which is mapping with the high quality to low quality. Third the user starts to train the neutral brainwave and any of commands (push, pull, left or right) brainwaves. Forth the user tries to use imagination command to move the virtual box.

The second, Experiments of BCI Training Methods: In order to enable the user wearing EPOC to control the viewport in 3D Max, we need to find out how to complete a training process to successfully zoom in/out of the virtual box via the imagination. Therefore the research firstly establishes an experiment procedure which is able to achieve over 80% accuracy to control the imagination command (which is valid signal). "Valid signal" means, in 5 seconds, after the voice command (e.g. pull), the user can successfully complete a 3D max command (zoom in the virtual box) via imagination through the EPOC device. EEG brainwaves is as the same as fingerprint, different users has their different pattern, in order to over 80% accuracy mind control, every user has to go through the training process before using the BCI embedded CAD system.

Therefore in order to find out a short and effective training process for every new user, we go through three different types of training procedure. In each training process, the user needs to achieve 80% accuracy to control the virtual object in 3D max. In each training experiment, the user will test the "imagination command" by randomly testing push or pull command for 50 times to evaluate the accuracy (see table 1).

Training experiment type 1 the user first goes through the neutral training 1 time and the push training 1 time and then try "imagination of push test" 10 time. And then repeat the previous action for 10 times. Second the user goes through the pull training 1 time and then tries the "imagination of pull test" 10 times. And then repeat the previous action for 10 times. Finally the user test the imaginations of push or pull for 50 times. The accuracy of imagination of push or pull mix command is 76%.

Training experiment type 2 the user goes through the each of neutral training, push training and pull training for 10 times. And then the user test the imaginations of push or pull for 50 times. The accuracy of imagination of push or pull mix command is 84%.

Training experiment type 3 the user goes through each of neutral training, push training and pull training for 20 times. And then the user test the imaginations of push or pull for 50 times. The accuracy of imagination of push or pull mix commands is 84%.

Table 1. BCI training experiments

Experiment type	Training procedure	Randomly test imagination of push or pull 50 times (Accuracy %)
Experiment type I	First step: Neutral training x1 time, Push training x1 time } Imagination of Push test x 10 times Second step: Repeat first action 10 times Third step: Pull training x1 time } Imagination of Pull test x 10 times Firth step: Repeat third action 10 times Fifth step: Randomly test imagination of push or pull 50 times	76%
Experiment type II	First step: Neutral training x10 times Push training x10 times Pull training x10 times Second step: Randomly test imagination of push or pull 50 times	84%
Experiment type III	First step: Neutral training x20 times Push training x20 times Pull training x20 times Second step: Randomly test imagination of push or pull 50 times	84%

In three different types of experiments, we can find that the imagination of push or pull test would not affect the result of accuracy. Therefore this research declared that the new user can achieve the 80% accuracy to use the mind command controlling the virtual box by going through the experiment type III.

5. SYSTEM IMPLEMENTATION

In the system framework, there are two building blocks, namely, 3ds Max plug-in and Emotiv API, respectively. The main task here is to provide an interface between 3ds Max plug-in and Emotiv API to realize the communication in one direction from EPOC+ hardware to 3ds Max software. Therefore, the requirement is just a single plug-in file in 3ds Max by using the Emotive API. Customizing our own plug-in function in 3ds Max is easy since the software itself is formed of many plugins. The most well-known part is the user interface around the 3ds Max which allows user to set their own way via Maxscript or dynamic link library. All the design of user-specified plug-in can accomplish by using 3ds Max SDK. Hereby, to our proposed system, the plug-in is event-triggered by the zoom in/out signals from EmoEngine, provided by Emotiv API and response zoom in/out of the object in 3ds Max. Keep reading the state of the user and

ISBN: 978-989-8533-38-8 © 2015

provide signal to judge whether the object should zoom in or out is the main program of our plug-in software. When one completes the training step, he or she can simply download the plug-in and start to customize the zoom in/out feature by imagination. That is, a simplest and most intuitive way.

6. SCENARIO DEMONSTRATION

In order to demonstrate the system, the subject goes through the training procedure. The user approximately used 18 minutes to finish the training procedure. In the scenario demonstration, the subject first wore the EPOC+ headband and checked all electrode contacts quality achieving green color. After the subject building a virtual box in 3Ds Max, he began to use the imagination command to “pull” the box. After 1.5 second the subject successfully zoom out the box of the viewport. And the subject keeps using mind command to zoom in the virtual box. After one second, the subject smoothly zooms in the box.

Figure 2. Scenario demonstration: A) user wears EPOC+, and check the electrode contact quality; B) the user started to use imagination command to zoom out the box of the viewport; C) the user smoothly zoom in the box of the viewport.

7. CONCLUSION

The research creates a “BCI embedded CAD” system prototype with the connecting with the connection between EPOC+ 3D studio Max via maxscript and C++. Through the system, the user is able to enhance the ability in 3D modeling through “thinking the commands”. For a new user, after going through the training procedure this research designed, he/she can finish the training process within 20 minutes. And the user is able to achieve the 80% accuracy using the mind command to push or pull the viewport in 3Ds Max.

However, in the scenario demonstration, sometimes the user used both “mind command” and “physical interface (keyboard or mice)” to control the virtual box. Therefore, the future study would focus on applying “BCI embedded CAD” system to cross 3D CAD platforms (3Ds Max, Rihno, and MAYA). It can be better to understand if the BCI embedded CAD users can perform higher efficiency than traditional users.

As to the contribution, the research would be significant not only in the architecture engineering but also the wide design fields. The limitation of the study, the system can only control on/off signal (pull or “invalid pull”) to zoom in/ out the viewport. In the future study, the system can be improved to precisely zoom in a bit, more, or partially viewport.

REFERENCES

Aish, R. (1979). 3d input for caad systems. Computer-Aided Design 11(2), 66–70.

Chapin, J. K., K. A. Moxon, R. S. Markowitz, and M. A. Nicolelis (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nature neuroscience 2(7), 664–670.

Huang, Y. C. (2011). How human-computer interface redefines original lifestyle in architecture? Advanced Materials Research 250, 1088–1097.

Krepki, R., B. Blankertz, G. Curio, and K.-R. Muller¨ (2007). The berlin brain-computer interface (bbci)– towards a new communication channel for online control in gaming applications. Multimedia Tools and Applications 33(1), 73–90.

Lichten, L. (1984). The emerging technology of cad/cam. In Proceedings of the 1984 annual conference of the ACM on The fifth generation challenge, pp. 236–241. ACM.

Sutherland, I. E. (1964). Sketch pad a man-machine graphical communication system. In Proceedings of the SHARE design automation workshop, pp. 6–329. ACM.

DEVELOPMENT OF MAINTENANCE-SUPPORTING IOT SYSTEM OF PLANT EQUIPMENT

Byung-Ju Lim, Chang-Dae Park and Kyung-Yul Chung
Environment and Energy Systems Research Division, Korea Institute of Machinery and Materials
156 Gajeongbuk-Ro, Yuseong-Gu, Daejeon, Korea (Republic of)

ABSTRACT

We developed the IoT(Internet of Things) systems for effectively increasing maintenance effectiveness of field-worker in plant. The system includes the movable or portable maintenance supporting equipment which can directly measure and analyze the various maintenance variables of equipment or system such as vibration, temperature, pressure, and etc. in the field, and the 3D technical information system which the workers can get the much information about the equipment such as location, specifications, maintenance procedures and design drawings. Our IoT system was installed in turbine floor of thermoelectric power plant in Korea and applied to maintain the related equipment.

KEYWORDS

Plant maintenance, Movable, Portable, IoT system

1. INTRODUCTION

In spite of expensing many worker and much costs for effectiveness maintenance of the plant, many shutdown events or safety accidents occur still [1]. To minimize these events, a supporting system to increase maintenance effectiveness of field-workers is needed.

In order that the field-workers get the information from drawings and documents of equipment, they have to look around the document storage places of the plant office, not easy to find them, very inconvenient. Also, in case of the equipment which could not get condition information such as temperature of equipment from online system, measurement system is needed to be separately installed in the field and additional cost and time is required. Especially, using additional measurement system, tiresome processes that treat the measured data and update to database server of plant are also needed. The information technology(IT) can solve easily the problem on these field-workers' inconveniences. IT can effectively reduce the unnecessary time and process between machines to humans [2]. Wireless network transport technology is possible to transport easily the various data related to equipment maintenance to the workers.

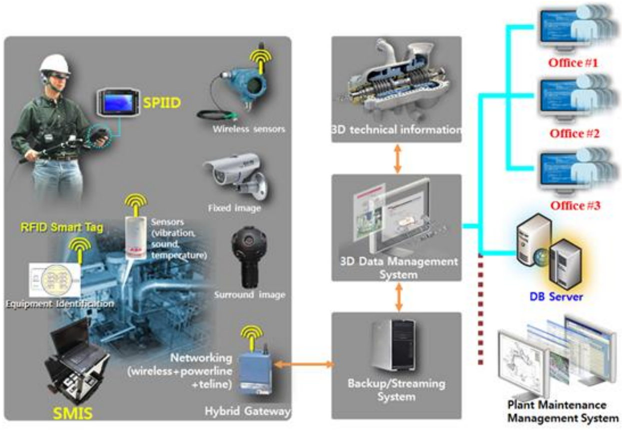


Figure 1. Schematics of maintenance-supporting IoT systems

ISBN: 978-989-8533-38-8 © 2015

(a) SMIS (b) SPIID

Figure 2. Smart maintenance-supporting equipment

Figure 3. 3D technical information system

In this study, we developed the IoT system for effectively increasing maintenance effectiveness of field-worker and then installed the system on thermoelectric power plant in Korea.

2. MAINTENANCE-SUPPORTING IOT SYSTEM

Our IoT system consists of hand-held device (SPIID, Smart Preventive Inspection and Information Device), movable cart(SMIS, Smart Maintenance Information System), various type of sensors, infrared camera, endoscope, 3D technical information system, 3D Data management system, data backup server and hybrid gateway as presented in Figure 1. The field-workers may use SPIID or SMIS to get the maintenance data from equipment, and are possible to get 3D information from the devices with field network.

Smart maintenance-supporting equipment consists of SPIID and SMIS as shown in Figure 2. SPIID as hand-held device is convenient to carry in the field and often uses the periodical maintenance or preventive maintenance which is relatively easy. But the numbers of data logging channels are less than four. SMIS looks like a luggage bag with wheels. Field-workers pull and install the SMIS near to equipment. SMIS is equipped with sensors and equipment to measure vibration, temperature, infrared-image and video image and can optionally applied to all the sensors with output signals of 0-10 Vdc and 4-20 mA. In step of failure initiation of equipment, SMIS is more suitable.

All plant equipment and system are converted into 3D model images as shown in Figure 3. In the GUI interface 3D technical information system, the workers can see as the same feature with maintenance equipment and decompose/assemble the 3D image to help to comprehend about structure and basic principle of operation of equipment.

3. APPLICATION AND CONCLUSION

The IoT system was installed in turbine floor of thermoelectric power plant in Korea and applied to maintain the related equipment. We installed the network using pager phone line which is already installed in plant and wireless network device. The SMIS is applied as maintenance supporting device near to turbine system. We monitored various condition signals of turbine with SMIS and get information such as 3D drawing, specifications, manuals and maintenance history related to turbine on GUI interface of SMIS.

We have developed the IoT(Internet of Things) systems for effectively increasing maintenance effectiveness of field-worker in plant. Installation of the system and connection to information was very convenient in the field.

REFERENCES

C. K. Woo. et al, 2014. Electricity-market price and nuclear power plant shutdown: Evidence from California. *Energy Policy*, Vol. 73, pp. 234-244.

K. K. Du. et al, 2013. Human machine interactive system on smart home of IoT. *J. China Univ. Posts Telecommun.*, Vol. 20, no. SUPPL. 1, pp. 96–99, 2013.

DEVELOPING OF EVALUATION METHOD OF MOTION REPEATABILITY BY MOTION CAPTURING

Macky Kato[1], Takeshi Sato[2] and Yoshie Shimodaira[3]
[1]*Waseda University, 279-15 Mikajima 2 Tokorozawa Saitama 359-1192 Japan*
[2]*Jissen Women's University, 1-1 Osakaue 4 Hino Tokyo 191-8510 Japan*
[3]*Nagano Prefectural College, 49-7 Miwa 8 Nagano 380-8525 Japan*

ABSTRACT

Motion repeatability is one of the required skills for any sports including tennis. One of the reasons why expert tennis players can accurately serve the ball to another side is their motion repeatability. On the other hand, however the novice players can luckily serve the ball, they are required to practice again and again to repeat the motion. The reliable criterion by the objective numerical indicator is needed for training and evaluation of motion repeatability. In this study, the method of numerical evaluation of motion repeatability would be developed by a motion capture system. The correlation coefficients of each segment angular fluctuation on the serving motions were validated as the criterion for motion repeatability. 17 sensors measure three-dimensional 23 segments and 22 joints angle fluctuations during the two seconds around the ball impact. Three experienced tennis players and three novice players tried to serve ten balls by right arms. The results of the experiments indicated that the repeatability of upper body tends to be stable comparing arms and legs for all the subjects. The correlation coefficients of the fluctuations at the measurement joints that are far from the upper body tend to be low. Comparing the experts and the novices on the motion of the arms and legs, the novices have much variation of correlation coefficients of each motion than the experts. In addition, most of the motion fluctuations of novices are larger than those of the experts. This study revealed that the joint angular fluctuation would be able to be applied as a criterion of motion repeatability.

KEYWORDS

Motion repeatability, Motion Capturing, Serving tennis balls, Joint angular fluctuation

1. INTRODUCTION

In the situation of task design, some of them require the skill to repeat their motion. For example, of small-scale task, keyboard texting and cursor pointing skill are required in an office job. On the other hand, motion in some sports requires almost all the muscle movement repeatability such as a golf swing (Evans 2012), tennis serving, and so on. The players need to train themselves to get repeatability, but many of the players cannot recognize their motion by themselves. In the situation of training, the sequence photographs and the video images are often used to evaluate the motion. It is not the objective evaluation, but just a subjective observation which depends on the personal feelings. Reliable evaluation should be installed as the objective numerical indicator, which indicates the repeatability of motion. The purpose of this study is developing the method of numerical evaluation of motion repeatability at the situation of serving a tennis ball. In the conventional studies, the optical motion capturing systems (MCS) are often installed to evaluate the motion (Schmitz 2014, Liu 2007). The experiments by the optical system need the lighting control and are limited to be done inside the room. However the ball games such as golf and tennis should be played in the air for duplicating their situation. In addition, the imagination of self-motion is not easy for the novices because there are too many parts for them to pay attention in the extremely short time. Though the novices' final goal is total motion repeatability, segmented motion evaluation would be effective as the first step. In this study, the MCS, which uses accelerometer and gyrometer, would be applied for quantification of the motion to evaluate the segmented motion repeatability.

ISBN: 978-989-8533-38-8 © 2015

2. METHOD

Three experienced tennis players and three novice players (Mean of Age: 20.8, Height: 170.2cm Weight: 61.6kg, right-handed) participated in the experiment. Each subject tried to repeat serving a ball into another side of the tennis court 10 times with the motion capturing system (Xsens MVN Awinda) and the same racket (Yonnex). Three-dimensional position fluctuations of lumbar vertebra were measured as a center of the body. In addition, 17 sensors shown in Table 1 measured three-dimensional angular velocities of 23 segments and 22 joints. The subject's motion would be reconstructed from the measured data to duplicate the motion as shown in Figure 1. Sampling rate was 40 fps. The fluctuations of two-second motion data around the ball impact, which is picked up and synchronized by cross-correlation analysis (MATLAB 2014b) from the measured time series are compared with other trials to build a correlation matrix of 10 time trials. The maximum correlation coefficient of three-dimensional value in the matrix would be used as the indicator of the repeatability of the serving motion. The effectiveness of the indicator would be verified by comparing the experienced players and the novice players.

Table 1. Sensor positions on the body

Segments	Sensor positions
Trunk	Head Chest Back
Arms	Shoulder (L/R) Upper (L/R) Forearm (L/R) Palm (L/R)
Legs	Thigh (L/R) Lower (L/R) Ankle (L/R)

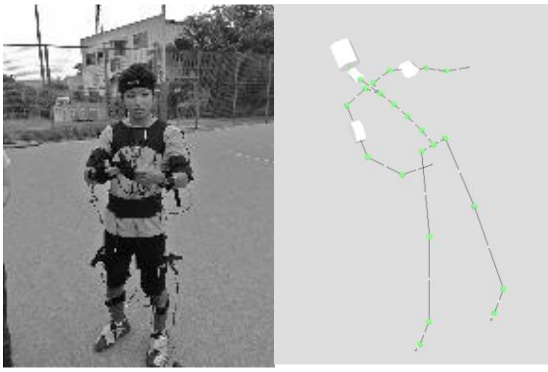


Figure 1. Motion Capturing System and captured body parts (MVN)

3. RESULT

Figure 2 shows the sample of the angular fluctuation of vertex and right wrist on the two-dimensional coordinates of x and y by an experienced player and a novice player. It shows that the difference between the angular fluctuations of the experienced player and those of the novice. Figure 3 shows the sample of maximum value of the correlation coefficient of three-dimensional angular fluctuation of vertex and right wrist. It shows that all the experienced players correlation is higher than the novices.

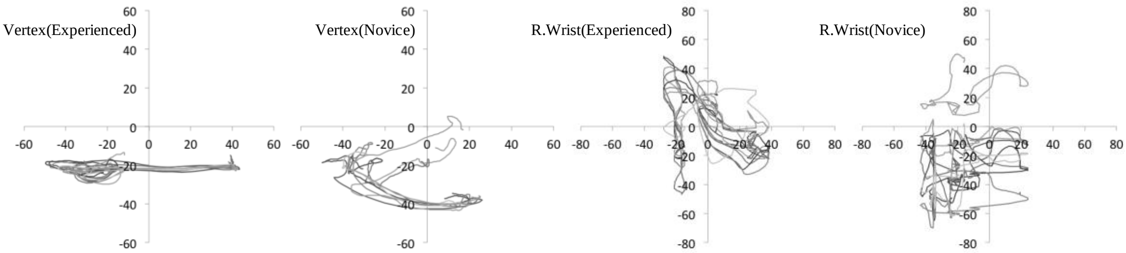


Figure 2. Vertex and Right wrist angular fluctuation of an experienced player and a novice player

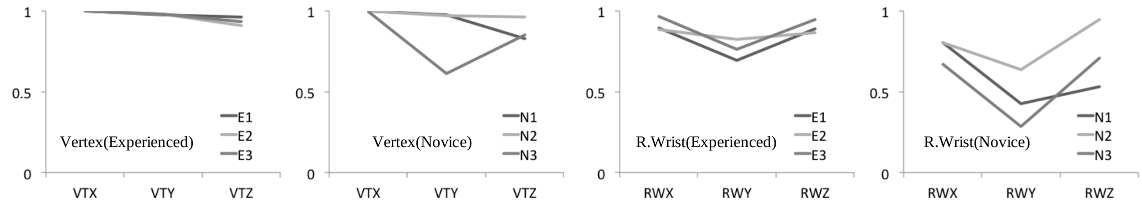
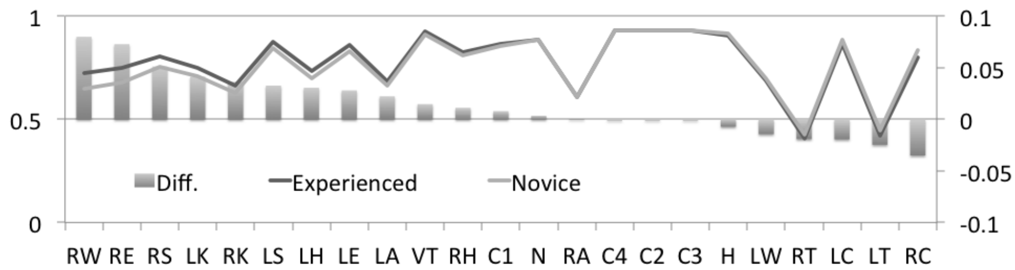


Figure 3. Correlation of Vertex and Right wrist angular fluctuation of an experienced player and a novice player

Comparing all the maximum of the correlation coefficient of x, y and z-axis of movement of lumber vertebra and 23 segments, experienced players' values were higher than the novices at 57% parts. Figure 3 shows the difference between the two groups of maximum correlation coefficient of three-dimensional value in the player's 10 trials. It shows the segments which have large differences between the two groups and those of small differences. Totally, there was a significant difference between two groups by paired t-test ($t=2.54$, $df=71$, $p<0.05$).



(VT: Vertebra, C1-C4;N;H: Upper Body, L/RC;L/RS;L/RE;L/RW: L/R Arm, L/RH;L/RK;L/RA;L/RT: L/R Leg)

Figure 4. Maximum value of correlation coefficient of each segment

4. DISCUSSION

Even though not all the joints show the difference of experience, the motion fluctuation in Figure 2 which means the stability of the upper body and wrist motion which is related to the direct racket control show the possibility to duplicate the accuracy of the ball serving. The correlation coefficients of some segments shown in Figure 3 would be able to be applied as the indicators of repeatability. In other words, the order of Figure 4 narrowed down the different joints in the motion between the experienced players and the novice. Especially the segments on the right arm, RW, RE and RS, seem to show the difference. In addition, both of the knee joint angular seem to be the same as the right arm. On the other hand, the differences of trunk movements, VT, C1, C2, C3 and C4 seem to be small. It can be said that they revealed the difference between the performance of experienced player and those of the novice. It can be said that the novice players will be able to pay attention to right arms through the training for motion repeatability by the feedback of the motion feature. The number of sensors will be decreased in the next step for the simplification of evaluation. It means the possibility to apply the motion repeatability evaluation system to the field.

REFERENCES

Kerrie Evans, et al., 2012, Repeatability of three-dimensional thorax and pelvis kinematics in the golf swing measured using a field-based motion capture system, Sports Biomechanics, Vol.11, No.2, pp.262-272

Anne Schmitz, et al., 2014, Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system, J. Biomechanics, Vol.47, 587-591

Helen Liu, et al., 2007. Accuracy and repeatability of an optical motion analysis system for measuring small deformations of biological tissues, J. Biomechanics Vol.40, pp.210-214

Doctoral Consortium

International Conferences Interfaces and Human Computer Interaction 2015,
Game and Entertainment Technologies 2015 and
Computer Graphics, Visualization, Computer Vision and Image Processing 2015

BINAURAL SOUND ANALYSIS AND SPATIAL LOCALIZATION FOR THE VISUALLY IMPAIRED PEOPLE

Oana Bălan, Alin Moldoveanu and Florica Moldoveanu
University Politehnica of Bucharest, Faculty of Automatic Control and Computers
Splaiul independenţei, 313, Bucharest, Romania

ABSTRACT

The blind people face serious difficulties concerning exclusion from working activities, lack of social involvement and having a sedentary lifestyle. The blind individuals can, however, enhance their life quality by using powerful and effective assistive devices that would help them to perform navigational and orientation tasks and to build a rich mental representation of the environment. An assistive system for the visually impaired people needs to fulfill certain usability requirements, such as to be wearable and affordable and to provide a large quantity of information as efficiently as possible in order to give the user a natural-like perception of the settings. The purpose of our research is to develop an assistive device for the blind people, based on alternative sensory modalities, such as hearing (by encoding the visual information into sound) and touch (by providing additional information through vibrations and other haptic cues). This paper presents the results achieved so far in our research, concerning the sonification and training techniques that will be applied for the development of the proposed assistive system.

KEYWORDS

3D sound, training, HRTF, virtual auditory display, haptic, blind people.

1. INTRODUCTION

In the scientific community, there have been significant attempts to develop information systems that would assist the blind people to accomplish different activities, such as navigating open or closed environments, gain spatial awareness or build a solid cognitive representation of the settings. There are several systems that enable the visually impaired people to use the computer, public transportation or to navigate in unknown environments. However, very few of them provide a high degree of independence and feasibility, as they fail in allowing the blind users to develop orientation and mobility skills and to navigate safely under unfamiliar circumstances [1] [2] [3]. Nowadays, due to the advances in computer technology, medicine and neuroscience, we believe that the development of a cost-effective, wearable and powerful assistive system that would replace sight by another sense can turn to be a realistic approach towards improving the life quality of the visually impaired people.

2. THE PROPOSED ASSISTIVE DEVICE

The system we intend to develop aims to provide a rich, continuous, real-time visual reconstruction of the environment that will be delivered to the user through spatialized 3D sounds and vibration cues. Thus, our device will offer a complex perception of the surrounding space, by emphasizing in particular the characteristics and dynamics of the objects and events that are of significant interest for the user. The data acquisition module will continuously scan the environment, acquiring real-time depth and color information that will be further processed, resulting in a set of relevant 3D objects. The next step consists in creating a spatial acoustic 3D model that would define the location and features of the target objects, based on a sonification technique that would employ 3D sounds, earcons, auditory icons, speech and by adjusting the physical parameters of the sound, such as pitch, amplitude and timings between two consecutive stimuli [4].

This research is oriented towards improving the blind people's spatial hearing resolution through perceptual feedback based training, multimodal sensorial adaptation (visual, auditory and haptic) and procedural learning [5] [6]. For this, we performed a series of experiments with both sighted and visually impaired subjects, in order to study and compare the level of spatial auditory performance they can achieve as a result of training. The theoretical and practical results obtained in my research will be used for creating an effective strategy for improving the sound localization accuracy of the visually impaired people (in respect with the development of a sensory substitution device aimed to provide a rich representation of the environment) and for designing audio games that are targeted for both the blind community and for the sighted players who want to try an alternative type of games [7] [8].

3. TRAINING THE SOUND LOCALIZATION SKILLS OF THE VISUALLY IMPAIRED PEOPLE

In this chapter, we intend to describe a strategy towards improving the sound localization accuracy of the blind and visually impaired people. This strategy is based on perceptual training and sensorial adaptation to 3D sounds synthesized with non-individualized Head Related Transfer Functions (HRTF). The HRTFs are a response that defines how the ear perceives a sound originating from a particular location in space [9]. They significantly depend on the anatomical features of the human body, such as size and shape of the pinna (the external ear), head and torso. These specific differences do not permit the use of the same HRTFs for all the listeners, as they would conduct to large localization errors and to an ambiguous acoustic perception [10]. Unfortunately, the use of individualized HRTFs in sound localization experiments or in virtual auditory environments is a very difficult task, due to the highly demanding measurement technique (it can take up to several hours to record the transfer characteristics for all the possible positions in the horizontal and in the vertical plane) that is tedious and tiring for both the subject and the experimenter [11]. Individualized HRTFs would conduct to a perfect virtual-reality auditory system but, as they are difficult to be obtained, most of the virtual auditory displays employ generic or non-individualized HRTF filters that lead to a decreased sound localization accuracy [12].

Our approach is built upon perceptual feedback based training, as a practical and effective solution towards improving the spatial auditory resolution in virtual acoustic systems and audio games. Thus, we developed a learning procedure aimed to help the visually impaired subjects to adapt to auditory cues that are different from their own by using haptic, visual (in the case of the subjects who have a higher degree of residual vision left) and acoustic feedback. Moreover, our training strategy intends to help the users recalibrate their spatial acoustic representation in unfamiliar virtual environments and to enhance their orientation and mobility abilities in navigational audio games or 3D sound based electronic travel systems. The proposed sound localization experiment has as main goal to improve the precision angular accuracy and to reduce the incidence of reversal errors (front-back and back-front confusions, situation in which the listener perceives the sound as originating from the opposite plane across the interaural axis – the axis that passes through the ears, dividing the horizontal space into the frontal and rear hemifields). As the rate of front-back confusions is relatively high when the listeners are exposed to 3D binaural sounds generated from non-individualized HRTFs, we considered devising a method for creating 3D audio stimuli that would enable the subjects to differentiate between the sources located in the front and in the back planes. Thus, we combined white and pink noise in varying proportions, according to the direction of the sound source in space, so that at 0 degrees to the front the listener perceived 100% white noise and 0% pink noise, at 90 degrees (to the left) and 270 degrees (to the right), the percentages of white and pink noise were equal (50%-50%), while at 180 degrees (to the back), the listeners could hear only pink noise. This spectral coloration of the sound helped the users to differentiate between the sound sources situated in the back and in the front, by creating a mental association (acquired as a result of the training process) between the spectral profile of the sound and the correct direction of its source.

Our experiment consisted of a pre-test, a training and a test session. In the pre-test and in the post-test sessions, the nine visually impaired listeners were required to listen to 3D sounds and to identify the location of a hidden target. The sonification technique was based on the inverse proportional encoding of distance, so that as the user got nearer to the target, the intensity of the sound increased and, on the other hand, when he got farther, the volume decreased until complete silence (outside the auditory area). During the training session, the subjects were asked to listen to a series of auditory stimuli and to indicate (using the conventional hour hand of the clock) the perceived direction of the sound. After the experimenter introduced the subjects' choice, they were offered instant haptic, auditory and visual feedback. The auditory feedback consisted in playing again the target sound over headphones.

On the other hand, the haptic feedback was provided through a series of vibrations on a haptic belt (that the listeners were required to wear on their heads) which were perceived from the same direction as the sound source. In addition to this, the visual feedback was offered graphically on the screen (for the subjects who had a higher degree of residual vision) – the subject's choice was colored in red, while the correct direction of the sound was colored in green (Fig. 1). In this way, the visually impaired subjects benefitted from various sensory perceptual experiences that allowed them to effectively create an association between the acoustic perception of the stimuli and the correct direction of the sound (Fig. 2).

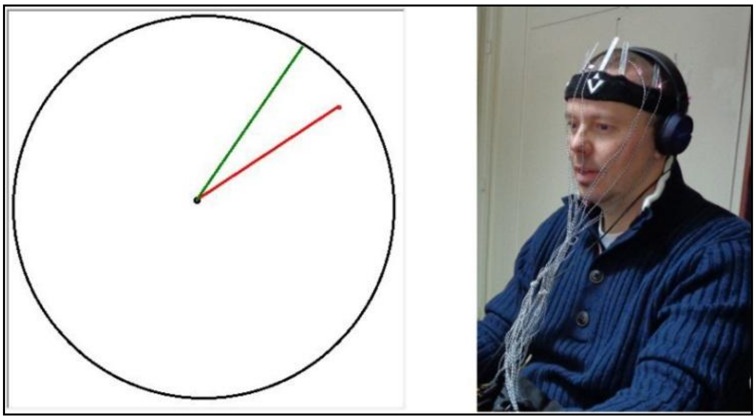


Figure 1. Visual (left) and haptic (right) feedback during the training session

Figure 2. Visually impaired subject and the experimenter during the haptic training session

The results of our experiment demonstrated that the visually impaired subjects succeeded to achieve a rapid adjustment of the auditory system in response to 3D sounds synthesized from non-individualized HRTFs, due to the crossmodal interaction between the auditory, visual and haptic senses. In addition to this, they improved their spatial auditory abilities, by reducing the front-back confusion rate and angular precision errors between the pre-test to the post-test session of the experiment, as a result of the training procedure. The adaptation process was rapid, demonstrating that the learning effect was enhanced by other cognitive mechanism, such as attention and focus. Moreover, the visually impaired users enhanced their spatial auditory resolution, navigational and orientation skills and decision-making abilities.

As a result, we can conclude that that training plays a fundamental role for auditory adaptation to altered hearing conditions and that a rapid adaptation of the auditory system to non-individualized HRTFs is possible through a spatial map recalibration with multimodal sensory associations. The proposed model can be considered an effective training tool for the future development of audio-only games or for the design of the assistive device aiming to provide a rich representation of the environment that will be delivered to the end-users through 3D binaural sounds and haptic cues.

4. CONCLUSIONS

To sum up, our paper briefly presented the results obtained so far in our research that is focused on developing an assistive device for the visually impaired people. Thus, we succeeded in developing a sound localization training strategy based on auditory and haptic feedback that significantly enhanced the spatial auditory resolution of the nine visually impaired subjects who participated in the experiment.

The results of the post-test session proved that the subjects have been able to use 3D binaural sounds as the only means for navigating in the virtual auditory environment. In addition to this, they improved their spatial auditory resolution, sound localization accuracy, orientation and mobility skills and directional decision-making abilities.

In conclusion, the proposed approach can be considered a useful training and rehabilitation tool for the future development of audio-only games or for the design of an assistive device for the blind people. Our research will continue by designing a navigational audio game (with hierarchical levels of difficulty), using the same sonification method as in the previous applications. Moreover, we intend to increase the duration of the training session and to make some non-invasive BCI EEG measurements for studying the subjects' neuroplasticity during the practice of the game. The results that will be obtained will be integrated into a training procedure for using the auditory sensory substitution device which will be developed in the next 2 years and for designing other audio games that will be available on an online platform dedicated to both the blind and the sighted players.

ACKNOWLEDGMENT

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/132395. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

REFERENCES

Balakrishnan G., Sainarayanan G., Nagarajan R., Yaacob S., 2007. Wearable Real-Time Stereo Vision for the Visually Impaired. Engineering Letters, 142 EL_14_2_2.

Bujacz M., Skulimowski P., Strumillo P., 2011. Sonification of 3D Scenes using Personalized spatial Audio to Aid Visually Impaired Persons. In Proceedings of the 17th International Conference on Auditory Display, 20-23 June 2011, Budapest, Hungary.

Dunai L., Fajarnes G. P., Praderas V. S., Garcia B. D., Lengua I. L., 2010. Real-time assistance prototype—A new navigation aid for blind people, Proc. IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society, 1173–1178.

Csapó Á., Wersényi Gy., 2013. Overview of auditory representations in human-machine interfaces. Journal ACM Computing Surveys (CSUR), Vol.46, No.2.

Bălan O., Moldoveanu A., Moldoveanu F., Negoi I., 2015. The Role of Perceptual Feedback Training on Sound Localization Accuracy in Audio Experiments. In Proceedings of The 11th International Scientific Conference eLearning and software for Education, Bucharest, April 23-24, 2015.

Mendonça C., 2014. A review on auditory space adaptations to altered head-related cues. Front Neurosci., Vol. 219, No. 8.

Bălan O., Moldoveanu A., Moldoveanu F., Dascălu M.I., 2014. Navigational 3D Audio-Based Game - Training Towards Rich Auditory Spatial Representation of the Environment. In Proceedings of the 18th International Conference on System Theory, Control and Computing, Sinaia, Romania, October 17-19, 2014, pp. 688-693.

Bălan O., Moldoveanu A., Moldoveanu F., 2015. Navigational Audio Games: An Effective Approach Towards Improving Spatial Contextual Learning for Blind People. The International Journal on Disability and Human Development, Volume 14, Issue 2, pp. 109-118, ISSN (Online) 2191-0367, ISSN (Print) 2191-1231, DOI: 10.1515/ijdhd-2014-0018, April 2015

Wersenyi Gy., 20112. Virtual Localization by Blind Persons. Journal of the Audio Engineering Society (AES), 60(7-8),pp. 579-586.

Dellepiane M., Pietroni N., Tsingos N., Asselot M., Scopigno R., 2008. Reconstructing head models from photographs for individualized 3D-audio processing. Comput. Graph. Forum, Vol. 27, No.7, pp. 1719-1727.

Meshram A., Mehra R., Yang H., Dunn E., Frahm J.M., Manocha D., 2014. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. International Symposium on Mixed and Augmented Reality.

Letowski T., Letowski S., 2011. Localization Error: Accuracy and Precision of Auditory Localization (Chapter 4) in Advances in Sound Localization, book edited by Pawel Strumillo, ISBN 978-953-307-224-1, published: April 11, 2011

ISBN: 978-989-8533-38-8 © 2015

SERIOUS GAME PATTERNS FOR SOCIAL INTERACTIONS

Barbara Reichart and Bernd Bruegge
Technische Universität München

ABSTRACT

Designing serious games for learning proves to be a very difficult task, as it requires knowledge not only about the content, but also about game design and pedagogy. There have been efforts to create collections of design patterns that help to create initial designs or improve existing ones. However those collections neglect the role of social interactions within serious games for learning, even though this could lead to improvements in learning. This is both due to technical and design reasons. In this paper we want to give an overview over social interaction patterns observed in existing serious games. They are meant as a starting point to introduce social interactions into serious games. One pattern is used as an example to demonstrate the patterns approach.

KEYWORDS

Serious game, pattern, learning, social interactions

1. INTRODUCTION

Serious games refer to games used for purposes other than entertainment. An important category are serious games for learning. Many researchers think that those can not only boost motivation, but also support creativity (MacDonald, Stodel, Farres, Breithaupt, & Gabriel, 2001), increase willingness to experiment, involve critical thinking and problem-solving, improve retention, and foster reflection. These hopes are founded in the fact that most games automatically fulfill basic requirements common to good learning environments (Norman, 1993).

While the expectations towards serious games for learning are high, it is not clear, how effective designs can be created. Achieving many, let alone all, of the goals listed above is extremely difficult. Besides creativity it requires deep understanding of educational theories and the subject matter of the game. On the one hand “a failure to base educational game design on well-established learning and instructional theories increases the risk of the game failing to meet its intended educational goals, and yielding students who are entertained, but who have not acquired any academic skills or knowledge.” (Gunter, Kenny, & Vick, 2007). On the other hand overlooking game design principles will lead to a product that trains the target skills, but has little in common with actual games.

In order to improve serious game design several researchers have developed pattern catalogues, which provide structured existing solutions on serious game design. While those catalogues provide a useful starting point, they rarely discuss the integration of social interactions into the game design. However integration of social interactions in a serious game for learning can prove useful for motivation, understanding, and retention.

In this paper we suggest several patterns, which are meant to give inspiration on different approaches on how social interactions can be integrated into serious game designs. We then describe two of them in detail. The patterns are gained from existing designs of serious games and based on other design pattern libraries.

2. SERIOUS GAME PATTERNS

Patterns are a general reusable solution to a commonly occurring problem. They have first been suggested by Alexander, who coined the term in the field of urban architecture (Alexander, 1987). Design patterns have later been picked up in the field of software engineering with huge success (Gamma, Helm, Johnson, & Vlissides, 2005). Authors like (Davidsson, Peitz, & Björk, 2004; Holopainen & Björk, 2003; Schell, 2008) have applied the approach to game design.

There already exist a few serious game pattern collections. All of them take slightly different approaches in their attempt to melt pedagogical and game play aspects. The Game Object Model (GOM) by (Amory & Seagram, n.d.; Amory, 2001, 2006) attempts to create a dialectic between pedagogical dimensions and game elements mainly through design patterns that incorporate the learning content within a **story**. Another approach focuses on generation of **flow** and **reflection** within problem-based gaming (K Kiili, 2005), (Kristian Kiili, 2007). It contains similar ideas as the previous model, but introduces the **cognitive load theory** as an important design factor in serious games. Lastly the Six Facets of Serious Games (Marne, Wisdom, Hyunh-Kim-Bang, & Labat, 2012; Marne, 2007) form a well-rounded design library, which concentrates on bringing together **ludic** and **pedagogical dimensions**.

3. PATTERNS FOR SOCIAL INTERACTIONS

3.1 Gap in Existing Pattern Catalogues

Existing pattern catalogues so far do not describe the consequences of social interactions within a game on the learning experience in detail. However according to the social development theory, learning is socially constructed (Vygotsky, 2007). If this is true, social interactions need to be considered during the design process. However due to heterogeneous learner groups, this can be difficult. This is why we suggest to extend the existing serious game pattern catalogues with patterns that show successful designs of social interactions in serious games.

3.2 Suggested Patterns

3.2.1 Overview

Social interaction in games can be grouped into two major categories: competition and cooperation. While both categories are not mutually exclusive (some games incorporate both cooperation and competition), those categories are still useful to identify and describe design patterns that can be used to support the respective categories.

Figure1

3.2.2 Identification Process

The patterns are extracted from existing games and serious games, as well as a literature research in the fields of serious games, pedagogy/psychology, educational technology, and game design patterns. Patterns were only accepted if at least three successful examples could be found.

ISBN: 978-989-8533-38-8 © 2015

3.2.3 Pattern Structure

Patterns are formulated in the pattern structure suggested by (Wellhausen & Fießer, 2011). It is based on the structured used by (Alexander, 1987), but extended by a several components. In the following we give an overview over the parts of the patterns.

- **Pattern Name**: A name to reference the pattern
- **Context**: The circumstances in which the problem is being solved
- **Problem**: explains the problem the pattern attempts to solve
- **Forces**: explains why the problem is difficult to solve
- **Solution**: explains the solution
- **Consequences**: gives an overview over benefits and liabilities of applying the pattern
- **Examples**: describes uses of the pattern in existing games and serious games

3.2.4 Example Pattern: Complimentary Roles

Context: You want to teach content that is too complex to keep all of it in mind at once and focusing on one aspect at a time is more promising (Kristian Kiili, 2007). Reflecting on different worldviews is an important goal (probably then competition is part of the equation, e. g. your common goal is to successfully lead a state, but each player leads their own department, with its own goals). Practicing cooperation is your main goal

♦♦♦

Problem: You want to create a game where the content is too complex for one person to keep all of it in mind at once. Hence you want to allow your players to focus on one aspect at a time (Kristian Kiili, 2007)**. Also you want to provide opportunities for discussion that can trigger reflection.**

Forces:

- It is ok to only have a rough understanding of the content that can be learned with the other roles.
- Observing other people in their roles already gives a good idea about the learning content.
- The content facilitates the creation of different roles. Roles do not feel artificial.
- People need to be able to play different roles.
- Probably requires a computer for every role.
- Implementing multi-player over network can be work intensive. Finding a way to make different roles work on one screen can be a tough design problem.
- The game needs to support communication between players of different roles.
- Creating a single player experience for the same game might be hard up to impossible.
- Shy players might have a hard time to join a team if they have to do so by themselves → provide automated matchmaking when possible
- Learners might prefer real life collaboration (Hummel et al., 2011)

♦♦♦

Solution: Provide the players with roles, which clearly differ in some regards. For example the roles could have different goals for success, a unique skill set, or access to information.

Benefit:

- Reduction of cognitive load to part of domain required to fill out the role
- Extermalization of knowledge schema provides opportunity for reflection.
- Replayability increases, as playing the game with a different role can provide another perspective.

Liability:

- Increase in cognitive load due to continuous need for communication
- Increase in development time and cost as each role requires additional resources
- Keeping game fair and balanced can be very hard

Examples:

The serious game 'Aquaculture' (Hummel et al., 2011) *allows players to take on the role of a project leader, who investigates and draws up a feasibility report on what would be the most suitable location to start a new shellfish production site. Two players play cooperatively one focusing on the ecological and the other focusing on governance perspective. In America's Army (www.americasarmy.com), a recruitment* *game for the US army, players take on different roles (rifleman, automatic rifleman, grenadier, and squad designated marksman), which differ in equipment and movement speed. This forces players to cooperate as a team.*

4. CONCLUSION

The pattern approach by no means guarantees an interesting and effective serious game for learning. However especially for beginners it can provide hints and suggestions for interesting design ideas they could pursue. The patterns suggested in this paper focused on the area of social interaction that so far has been overlooked by many serious games for learning. They are based on existing serious games, however as the numbers of those games is limited, the suggested patterns are probably incomplete. Further evaluation of the patterns is required to demonstrate not only that they are used, but also that they improve existing designs.

REFERENCES

Alexander, C. (1987). *A Pattern Language: Towns, Buildings, Construction*. Oxford University Press. Retrieved from http://www.amazon.de/Pattern-Language-Buildings-Construction-Environmental/dp/0195019199

Amory, A. (2001). Building an Educational Adventure Game: Theory, Design and Lessons. *Journal Of Interactive Learning Research*, *12*(2/3), 249–263. Retrieved from http://www.questia.com/PM.qst?a=o&se=gglsc&d=5002415705

Amory, A. (2006). Game object model version II: a theoretical framework for educational game development. *Educational Technology Research and Development*, *55*(1), 51–77. doi:10.1007/s11423-006-9001-x

Amory, A., & Seagram, R. (n.d.). Educational Game Models : Conceptualization and Evaluation Educational Game Models : Conceptualization and Evaluation. *Journal of Higher Education*, *27*(0), 206 – 217.

Davidsson, O., Peitz, J., & Björk, S. (2004). *Game Design Patterns for Mobile Games*.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (2005). *Design Patterns: Elements of Reusable Object-Oriented Software with Applying UML and Patterns: an Introduction to Object-Oriented Analysis and Design and Iterative Development*. Addison Wesley. Retrieved from http://www.amazon.com/Design-Patterns-Object-Oriented-Introduction-Development/dp/1405837306

Gunter, G. a., Kenny, R. F., & Vick, E. H. (2007). Taking educational games seriously: using the RETAIN model to design endogenous fantasy into standalone educational games. *Educational Technology Research and Development*, *56*(5-6), 511–537. doi:10.1007/s11423-007-9073-2

Holopainen, J., & Björk, S. (2003). Game Design Patterns.

Hummel, H. G. K., van Houcke, J., Nadolski, R. J., van der Hiele, T., Kurvers, H., & Löhr, A. (2011). Scripted collaboration in serious gaming for complex learning: Effects of multiple perspectives when acquiring water management skills. *British Journal of Educational Technology*, *42*(6), 1029–1041. doi:10.1111/j.1467-8535.2010.01122.x

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education*, *8*(1), 13–24. doi:10.1016/j.iheduc.2004.12.001

Kiili, K. (2007). Foundation for problem-based gaming. *British Journal of Educational Technology*, *38*(3), 394–404. doi:10.1111/j.1467-8535.2007.00704.x

MacDonald, C. J., Stodel, E. J., Farres, L. G., Breithaupt, K., & Gabriel, M. A. (2001). The demand-driven learning model. *The Internet and Higher Education*, *4*(1), 9–30. doi:10.1016/S1096-7516(01)00045-8

Marne, B. (2007). *The Six Facets of Serious Games How to mix fun and learning in serious games ?*

Marne, B., Wisdom, J., Hyunh-Kim-Bang, B., & Labat, J.-M. (2012). The Six Facets of Serious Game Design: a Methodology Enhanced by our Design Pattern Library, 1–14.

Norman, D. A. (1993). *Things that make us smart - Defending human attributes in the age of the machine.* Reading, MA: Addison-Wesley.

Schell, J. (2008). *The Art of Game Design.* Elsevier.

Vygotsky, L. (2007). Does Vygotsky hold the key to reinvigorating the modern education system ? *Children*, (August).

Wellhausen, T., & Fießer, A. (2011). How to write a pattern. *European Conference on Pattern Languages of Programs, EuroPloP, 11*.

AUTHOR INDEX